98-03

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

STOCHASTIC PERMANENT BREAKS

BY

ROBERT F. ENGLE

AND

AARON D. SMITH

DISCUSSION PAPER 98-03
JANUARY 1998

STOCHASTIC PERMANENT BREAKS

BY ROBERT F. ENGLE AND AARON D. SMITH[1]

This paper aims to bridge the gap between processes where shocks are permanent and those with transitory shocks by formulating a process in which the long run impact of each innovation is time varying and stochastic. Frequent transitory shocks are supplemented by occasional permanent shifts. The stochastic permanent breaks (STOPBREAK) process is based on the premise that a shock is more likely to be permanent if it is large than if it is small. This formulation is motivated by a class of processes that undergo random structural breaks. Consistency and asymptotic normality of quasi maximum likelihood estimates is established and locally best hypothesis tests of the null of a random walk are developed. The model is applied to relative prices of pairs of stocks and significant test statistics result.

KEYWORDS: Structural breaks, nonlinear moving average, unit roots, quasi maximum likelihood estimation, Neyman-Pearson testing, locally best test, temporary cointegration.

1. INTRODUCTION

Time series analysts tend to draw a sharp line between processes where shocks have a permanent effect and those where they do not. The most notable example of this is the distinction between stationary AR(1) processes, where all shocks are transitory, and the random walk. As the autoregressive root approaches one, the rate at which shocks are expected to decay decreases, but they remain transitory. This paper aims to bridge the gap between transience and permanence by formulating a process in which the long run impact of each observation is time varying and stochastic. At one extreme all innovations are transitory and at the other, all shocks are permanent.

The concept of varying the permanent impact of shocks is linked to the familiar topic of structural change. Whenever a shock, or part of a shock, has a permanent effect we can interpret this as a specific type of structural break. Under this definition, a random walk has a break every period but a stationary ARMA process has no breaks. Processes such as a threshold autoregressive (Tong (1983)) have no breaks. The parameters change values, which causes the innovations decay at a different rate, but nonetheless they remain transitory.

The stochastic permanent breaks (STOPBREAK) process is motivated by a class of processes that incur random structural shifts at random intervals. Analysis of a structural shift when the break point is known a priori generally involves standard test statistics and estimators (Chow (1960)). When the break point is unknown the problem becomes more difficult because the break point must be estimated and, under the null of no break, this parameter is unidentified. Andrews, Lee and Ploberger (1996), Andrews (1993), Hansen (1992), Christiano (1992) and others have studied this problem in various contexts. When considering multiple break points the problem becomes further complicated since it requires specification of the number of breaks and inclusion of enough parameters to account for each regime. This becomes intractable when the number of break points becomes large.

From a forecasting perspective, the errors of finding too many breaks and not enough breaks are very different, resulting in either bias or imprecision. Linear moving average smoothers often forecast this type of data relatively well, but lack flexibility. We approach the problem from a different angle, treating the breaks endogenously by inferring their magnitude and frequency from realizations on a single random variable.

This allows the model enough flexibility to react to breaks without overloading on parameters.

We apply the STOPBREAK model to relative prices of pairs of stocks, conjecturing that a pair of stock prices may move together for periods of time and jump apart occasionally. They may exhibit a type of temporary cointegration. We expect this relationship to be strongest between stocks within the same industry, as they are likely to have more common components in their stock price determinants.

The paper proceeds as follows. In the next section, we introduce the process and discuss its properties and its relation to other non linear time series processes. Sections 3 and 4 treat hypothesis testing and Section 5 estimation issues. Empirical results follow in Section 6.

2. STOPBREAK PROCESS

In its simplest form, the STOPBREAK process is:

(1) $$y_t = m_t + \varepsilon_t \qquad (t = 0,1,\ldots,T).$$

where $m_t = E(y_t | \Im_{t-1})$ is a time varying conditional mean which is updated via

(2) $$m_t \equiv m_{t-1} + q_{t-1}\, \varepsilon_{t-1}$$

$$= m_0 + \sum_{i=1}^{t} q_{t-i}\varepsilon_{t-i} \qquad (t = 1,2,\ldots,T).$$

We assume that m_0 is fixed and known and that $\{\varepsilon_t, \Im_t\}$ is a martingale difference sequence, where $\{\Im_t\}$ denotes an increasing sequence of σ-fields. The function $q_t = q(\varepsilon_t)$ is bounded by zero and one and defined such that $E(q_t\varepsilon_t | \Im_{t-1}) = 0$ and

$$\frac{\partial q}{\partial \varepsilon}\bigg|\left|\varepsilon_t\right|$$

is non-negative and finite[2] wp 1.

This formulation is based on the assumption that the time series is less likely to mean revert after a large shock than after a small one. If $\tilde{q}_t$ = 1, then the realized process at time t is a random walk[3]. If $\tilde{q}_t$ = 0, the conditional mean does not change and consequently neither does the long run forecast for y_t. Thus we have a process where the permanence of a shock is determined endogenously. For example, in the stock market, investors may perceive large shocks as containing significant informational content and small shocks as mere noise. Consequently, their valuations and expectations only react to large shocks. Biological systems may fluctuate around some constant level of fitness with occasional mutations having large permanent effects. An economy may be subject to sporadic permanent supply shocks and frequent transitory demand shocks. Technology growth and crime rates are examples of other series that could potentially behave similarly.

Identification of permanent shocks by their magnitude should be viewed as a special case of this process. In general, any factor that is part of the information set could be an argument in the function q_t. For example, if modeling stock prices, relevant variables may include macroeconomic announcements, profit announcements, interest rates, exchange rates etc. However it is unlikely that one could account for all potential factors that could cause a permanent shift and even if it were possible, it would overload the process with parameters.

In this paper, we take an agnostic approach, assuming only that permanent shifts will largely be reflected in an innovation that is larger than the norm. A model specified in this manner may not pick up small shifts in the mean and thus will subsequently make systematic errors until the mean moves sufficiently. It will also be prone to overreacting to large transitory shocks, indicating equal and opposite breaks rather than no break. Nonetheless we maintain that the simplicity and flexibility of the process more than offset these negatives if the goal is to obtain conditional forecasts.

The STOPBREAK process can also be considered a type of error correction mechanism in the sense that it is reactionary, rather than anticipatory. It makes no attempt to predict when a permanent innovation will occur, but merely reacts to shocks by forecasting their degree of permanence.

2.1. *Relation to Other Time Series Processes*

The distinguishing feature of the STOPBREAK process is that the permanent effect of shocks is time varying and stochastic. In some periods, breaks are permanent and in other periods they are not.

DEFINITION: Consider some stochastic process $\{y_t\}$, t=0,1, ..., T. The *permanent effect of observation t* is

$$\lambda_t \overset{d}{=} \lim_{k\to\infty} \frac{\partial f(y_t, k)}{\partial y_t},$$

where $f(y_t, k) \equiv E(y_{t+k} | y^t)$, y^t denotes the entire past history of $\{y_t\}$ and $\overset{d}{=}$ signifies equality in distribution.

A martingale has the property that all shocks have a permanent effect, i.e. λ_t = 1 wp 1 $\forall t$. Conversely, if λ_t = 0 wp 1 $\forall t$, the process has no permanent breaks. Realizations, $\tilde{\lambda}_t$, between zero and one indicate partial permanent breaks, so that some fraction of a shock is remembered. An example is an integrated process with a negative invertible moving average component. If $\tilde{\lambda}_t < 0$, we have negative permanent breaking where the process over corrects for shocks, and if $\tilde{\lambda}_t > 1$, shocks are magnified, i.e. the permanent effect is greater than the initial effect. A process with positively correlated first differences would have $\lambda_t > 1$ wp 1 $\forall t$. A permanent break is deemed to have occurred at time *t* if the realized permanent impact of that observation is non-zero i.e. if $\tilde{\lambda}_t \neq 0$.

A *k* period ahead forecast of the STOPBREAK process in (2) is:

$$E\left(y_{t+k} \middle| y^t\right) = m_{t+1} \equiv m_t + q_t \varepsilon_t$$

since $E\left(m_{t+k} \middle| y^t\right) = m_{t+1}$. Differentiating with respect to ε_t gives the permanent impact of observation t as

$$\begin{aligned} \hat{\lambda}_t &= q_t + \left.\frac{\partial q}{\partial \varepsilon}\right|_{\varepsilon_t} \varepsilon_t \\ &= q_t\left(1 + \eta_{q,t}\right). \end{aligned}$$

Since both q_t and $\eta_{q,t}$ are non-negative wp 1, we have $\lambda_t \geq 0$ wp 1 $\forall t$. In the STOPBREAK process, the long run impact of shocks varies over time.[4]

Compare this to a stationary autoregressive process in which the coefficients change values, i.e. $y_t = \rho_t y_{t-1} + \varepsilon_t$ where ε_t is zero mean i.i.d. and ρ_t is a random variable taking the values ρ_1 and ρ_2 each with positive probablility. The mechanism that determines ρ_t

could, for example, be governed by a markov chain as in Hamilton (1989) or a threshold (Tong (1983)). As long as the process is stationary in all regimes, it will have $\lambda_t = 0$ wp 1 $\forall t$. The rate at which shocks decay changes across regimes, but the time series remains fully mean reverting. In fact, even if, say ρ_1, is unity the permanent effect of all observations will still be zero with probability one. This arises because, assuming that each regime is realized with positive probability, there will eventually be enough periods of stationarity for the effect of a shock to disappear.

The stochastic unit root process of Granger and Swanson (1997), where ρ_t varies stochastically around one, is also an interesting case. Consider $\rho_t = exp(\nu_t)$ where ν_t is a stationary Gaussian series with mean μ_ν and variance σ_ν^2. If ν_t contains some positive temporal correlation and $E(exp(\nu_t)) = 1$, then the effect of past shocks is magnified and λ_t is infinite wp 1. If ν_t is such that $E\left(exp\left(\sum_{k=1}^{\infty} \upsilon_{t+k}\right)\right) < 1$, all shocks eventually die away, i.e. $\lambda_t = 0$ wp 1 $\forall t$. Between these two extremes is a knife edge where the process exhibits stochastic permanent breaks. However, the permanent impact of shocks tends to fluctuate around one in this case, so the process does not bridge the gap between permanence and transience as STOPBREAK does.

Consider the process:

$$y_t = m_t + \varepsilon_t \qquad (t = 1,2,\ldots,T) \tag{3}$$

where $m_t = m_{t-1} + u_t$, $prob(u_t{=}0) = (1-p_u)$, $prob(u_t \sim N(0,\ \sigma_u^2)) = p_u$, m_0 is fixed and known and $\varepsilon_t \sim N(0,\ \sigma_\varepsilon^2)$. We can rewrite (3) as

$$\Delta y_t = u_t + \varepsilon_t - \varepsilon_{t-1}.$$

When $\tilde{u}_t = 0$, the innovation in period t is transitory, i.e. $\tilde{\lambda}_t = 0$. Otherwise the u_t component of the period t innovation is permanent i.e. we have a partial permanent break. Thus this process is similar to STOPBREAK in the sense that innovations vary stochastically between permanence and transience.

The best linear representation for the process in (3) is an integrated moving average or, by another name, the exponential smoother. The exponential smoother has constant partial permanent breaks wp 1 in each period. A fixed proportion of each shock remains permanent, where this proportion is determined by the probability of a permanent shock, p_u, and by the relative variances of the two innovation terms. We show in Section 6 that a STOPBREAK model is able to forecast the process in (3) significantly better than the exponential smoother.

2.2. *Properties of the STOPBREAK Process*

We can rewrite the basic STOPBREAK in (1) and (2) process as

$$(4) \qquad \Delta y_t = \varepsilon_t - \theta_{t-1}\varepsilon_{t-1} \qquad (t = 1,2,\ldots,T)$$

where $\theta_{t-1} = 1 - q_{t-1}$. To forecast from this process, we require an estimate of ε_{t-1} i.e. we require the process to be invertible.

THEOREM 1: *The non-linear moving average process in (4) is invertible wp* 1 *if* *prob*($q_t > 0) > 0$ *and if* $q_t(1+\eta_{qt}) < 2$ *wp* 1 $\forall t$, *where* $\eta_{qt} = \frac{\varepsilon_t}{q_t}\frac{\partial q}{\partial \varepsilon}\Big|_{\varepsilon_t}$.

PROOF: See Appendix.

Invoking the conditions of Theorem 1, we have $\lambda_t < 2$ wp 1 $\forall t$. However intuition suggests that the majority of the probability mass for λ_t would lie in the [0,1] interval. Values between one and two may arise from second order effects through the function q_t. An increase in ε_t raises the long run forecast of y_{t+k} firstly by proportion q_t and secondly by raising the value of q_t. Thus, depending on the properties of the function q_t, the final effect of a change in ε_t may exceed the value of the change i.e. the permanent effect of shocks may be greater than one.

In choosing a functional form for q_t, we restrict our attention to functions that satisfy the conditions in Theorem 1. The simplest example is a threshold function, where q_t takes the value one for values of ε_t greater in absolute value than some threshold, and zero otherwise. Although the elasticity of q_t with respect to ε_t is infinite at the threshold, we can still claim invertibility wp 1 since this event occurs on a set of measure zero.

We choose to specify q_t as a continuous function. This allows for partial permanent breaks in the process, so that shocks in the gray area between large and small have only some proportion that is permanent. Intuition suggests that this less rigid specification may better represent much empirical data.

Suppose that a correctly specified model for q_t is:

$$(5) \qquad q_t(\gamma) = \frac{\varepsilon_t^2}{\gamma + \varepsilon_t^2}\ , \quad \gamma > 0.$$

Under this specification, there is a smooth transition between the two extremes as shown in Figure 1. A more rapid transition is obtained by increasing the exponent on ε_t to four. Increasing it further to six leads to a violation of the conditions in Theorem 1.

Along with its simplicity, an advantage of this specification is that it collapses to one as γ goes to zero. This enables parametric testing of the random walk null hypothesis. A number of other specifications, such as the logistic, can only collapse in this fashion if an extra parameter is added. Then, under the null there is an unidentified parameter, which introduces further complications to the testing problem.

Another useful property of the specification in (5) is that, for all $\gamma > 0$, $q_t(\gamma) = 0$ if and only if $\varepsilon_t = 0$ i.e. no non-zero shocks are completely transitory. Thus in periods of small errors the process is only approximately stationary. This approximation proves beneficial for hypothesis testing since it yields test statistics with the standard distributions. We return to this topic in Section 3.

2.3. *Generalizations*

The STOPBREAK process is a special case of the following general breaking process:

(6) $$A(L)\,B(L)\left(y_t - x_t'\delta\right) = z_{t-1}A(L)\varepsilon_t + \left(1 - z_{t-1}\right)B(L)\varepsilon_t \qquad (t = 1,2,\ldots,T)$$

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$$

$$B(L) = 1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_s L^s$$

where x_t denotes a vector of explanatory variables, ε_t an innovation term, z_{t-1} some measurable function of information up to t–1, and L the lag operator. When z_{t-1}=0 wp 1 $\forall t$, $B(L)$ is a common autoregressive and moving average factor which cancels out to leave $A(L)y_t = \varepsilon_t$. Similarly, the process reduces to $B(L)y_t = \varepsilon_t$ when z_{t-1}=1 wp 1 $\forall t$.

Setting $\delta=0$, $B(L) = 1-L$, $A(L) = 1$, and $z_{t-1}=q_{t-1}(\gamma_0)$ obtains the basic STOPBREAK process.

The general process in (6) will only exhibit a changing permanent effect of shocks if one of the lag polynomials, say $B(L)$, contains a unit root and if the other one has all roots outside the unit circle. This causes the effect of innovations to range between permanent and transitory. For example, consider $\delta=0$, $B(L) = 1-L$, $A(L) = 1$ and $z_{t-1}=1$ wp 1 if $t=t^*$ and $z_{t-1}=0$ wp 1 if $t \neq t^*$. This process has a break in its mean at $t = t^*$, i.e. $\lambda_t=1$ wp 1 at $t=t^*$ and $\lambda_t=0$ wp 1 in all other periods.

The general formulation in (6) also reveals a number of possible generalizations to the simple STOPBREAK process. For example, the process could have some temporal correlation when in 'non-breaking' periods. This corresponds to $\delta=0$, $B(L) = 1-L$, $A(L) = 1-\alpha_0 L$, and $z_{t-1}=q_{t-1}(\gamma_0)$, and implies the moving average representation:

$$\text{(7)} \qquad \Delta y_t = \alpha_0 \Delta y_{t-1} + \varepsilon_t - \theta_{t-1}\varepsilon_{t-1} \qquad (t = 1,2,\ldots,T)$$

where $\theta_{t-1} = 1-(1-\alpha_0)q_{t-1}(\gamma_0)$ and $0 \leq \alpha_0 < 1$. Now y_t has an AR(1) and a random walk as its two extremes.

Including explanatory variables implies a type of temporary cointegration since it implies that a linear combination of variables follows a STOPBREAK process. This linear combination is approximately stationary for periods of time before moving and then remaining nearly stationary at a new level for a period of time. The cointegrating coefficients do not change. This parallels the practice of intercept correction, which is often used in forecasting (see Hendry and Clements (1996)). The intercept is allowed to shift to correct for mean shifts while keeping the fundamental relationship between the variables constant.

There are other special cases of (6) where λ_t is equal to some constant wp 1. For example, if $\delta=0$, $A(L) = 1-\alpha L$, $B(L) = 1-\beta L$, and if $z_{t-1} = 0$ wp 1 $\forall\ t < t^*$ and $z_{t-1} = 0$ wp 1 $\forall\ t \geq t^*$, we have an AR(1) process where the autoregressive parameter transforms from α to β at $t=t^*$. This formulation is slightly different from a conventional parameter shift, since the break in this case is not clean. Unless $\tilde{y}_t$ is equal to its unconditional mean at the change point, the influence of that point decays away exponentially.

3. HYPOTHESIS TESTING

In the long run, a STOPBREAK random variable has no tendency to return to any previous point. With probability one, there will be a period with a non-zero permanent break, after which point the past has no predictability. Thus the process is not covariance-stationary and its spectral density at frequency zero is infinite. The long run properties of the series are like those of a random walk.

THEOREM 2: *Suppose that* $\{y_t\}$, $t=0,1\ldots,T$, *is a stochastic process represented by (1) and (2) with* $q_t(\gamma_0)=\varepsilon_t^2/(\gamma_0+\varepsilon_t^2)$, $0 < \gamma < \infty$. *Let* $\{\varepsilon_t, \Im_t\}$ *be a strictly stationary ergodic martingale difference sequence such that* $\varepsilon_t|\Im_{t-1}$ *is symmetrically distributed. Then*

$$T\hat{\rho} = \frac{T^{-1}\sum_{t=1}^{T}\Delta y_t y_{t-1}}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{d} \frac{\sigma_0\int(\sigma_0 W - \gamma_0\omega_0 W_\gamma)dW - \gamma_0\omega_0\int(\sigma_0 W - \gamma_0\omega_0 W_\gamma)dW_\gamma}{\int(\sigma_0 W - \gamma_0\omega_0 W_\gamma)^2}$$

where W and W_γ are standard Brownian motions on the unit interval, σ_0^2 *=* $E(\varepsilon_t^2) < \infty$ *and* $\omega_0^2 = E\left(\varepsilon_t^2 / \left(\gamma_0 + \varepsilon_t^2\right)^2\right)$.

PROOF: See Appendix.

The implication of Theorem 2 is that, even in large samples, Dickey Fuller (DF) type tests never reject the unit root null with probability one, when the data is generated from a STOPBREAK process. However there is some power because the downward bias in $\hat{\rho}$ is increasing in γ_0. Table I lists the simulated power of a simple DF test against γ_0/σ_0^2 for a sample size of 1000 using the specification in (7).

The inconsistency of the DF test arises due to the long run similarity between the STOPBREAK and a random walk.[5] Similar logic implies that unit root tests which have stationarity as the null (e.g. Saikkonen and Luukkonen (1993)) will have power against a STOPBREAK process. Thus, specific tests to distinguish between a random walk and a STOPBREAK process are desirable. Essentially these tests will search for plateaus in y_t or, more specifically, periods where the permanent effect of shocks is low.

Using the parameterization in (5), we can write a model for the process in (1) and (2) as

$$\Delta y_t = \frac{-\gamma \varepsilon_{t-1}}{\gamma + \varepsilon_{t-1}^2} + \varepsilon_t .$$

From this we see that the random walk null can be formulated parametrically as a test of H_0: $\gamma = 0$. Consider testing against the point alternative $\gamma = \bar{\gamma}$. From the Neyman-Pearson lemma, the most powerful test rejects for large values of the likelihood ratio. Given

Theorem 1, we can form the Gaussian conditional log likelihood functions under the null and under the alternative, and write their difference as

$$L_1 - L_0 = -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\left(\Delta y_t + \frac{\bar{\gamma}\varepsilon_{t-1}}{\bar{\gamma}+\varepsilon_{t-1}^2}\right)^2 - (\Delta y_t)^2\right)$$

(8) $$= -\frac{\bar{\gamma}}{\sigma^2}\sum_{t=1}^{T}\left(\Delta y_t \frac{\varepsilon_{t-1}}{\bar{\gamma}+\varepsilon_{t-1}^2}\right) - \frac{\bar{\gamma}^2}{2\sigma^2}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}}{\bar{\gamma}+\varepsilon_{t-1}^2}\right)^2 .$$

Thus the most powerful test of the random walk null is a linear combination of two statistics, each of which depend on the values of the parameters under the alternative. It follows that there is no uniformly most powerful test of a random walk against a STOPBREAK model.

THEOREM 3: *Assume that ε_t is a strictly stationary $\Im_t$-measurable random variable with uniform mixing coefficient $\phi(j) = o(j^{-\kappa})$ where*

$$\kappa = \left(\frac{1+\eta}{\eta}\right)\left(\frac{\xi}{1/4-\xi}\right)$$

for $\eta > 0$ *and* $0 < \xi < 1/4$ *defined such that*

$$E\left(\frac{|\varepsilon_t|^{1+\eta}}{\left(c/\sqrt{T}+\varepsilon_t^2\right)^{1+\eta}}\right) = O\left(T^{\xi(1+\eta)}\right).$$

Then, for $c>0$,

$$\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t^2}{\left(c/\sqrt{T}+\varepsilon_t^2\right)^2}\right) - E\left(\frac{\varepsilon_t^2}{\left(c/\sqrt{T}+\varepsilon_t^2\right)^2}\right)\right] \xrightarrow{L_1} 0 ,$$

where $\xrightarrow{L_1}$ *denotes convergence in* L_1 *norm.*

PROOF: See Appendix.

Suppose that $\bar{\gamma}$ is local to zero, i.e. $\bar{\gamma} = \bar{c}/\sqrt{T}$. Then, from Theorem 3, the second term in (8) is asymptotically equivalent to a constant. This implies that

$$\sum_{t=1}^{T}\left(\Delta y_t \frac{\varepsilon_{t-1}}{\bar{\gamma}+\varepsilon_{t-1}^2} \right)$$

is a sufficient statistic for the locally best test, which suggests using a t-test of H_0: ϕ=0 against a negative alternative in the following regression:

(9) $$\Delta y_t = \varphi \frac{\Delta y_{t-1}}{\bar{\gamma}+\Delta y_{t-1}^2} + u_t .$$

The standard distribution theory applies to this t-statistic. This arises because the form of q_{t-1} means that the process is never exactly stationary, implying that, under both the null and the alternative, Δy_t contains no unit moving average roots. The result is formalized below under general assumptions on ε_t.

THEOREM 4: *Suppose that* $\{y_t\}$, t=0,1…,T, *is a stochastic process represented by (1) and (2) with* $q_t(c_0) = \varepsilon_t^2 / \left(c_0/\sqrt{T} + \varepsilon_t^2\right)$, c_0>0. *Define*

$$\omega_{T\gamma\bar{\gamma}} = E\left(\frac{\varepsilon_t^2}{\left(c_0/\sqrt{T}+\varepsilon_t^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_t^2\right)} \right),$$

$$\omega_{T\sigma\bar{\gamma}} = E\left(\frac{\varepsilon_t^2 \varepsilon_{t-1}^2}{\left(\bar{c}/\sqrt{T}+\varepsilon_t^2\right)^2} \right),$$

$$Y_{Ti} = \sum_{t=(i-1)b_T+l_T}^{ib_T} \left(\frac{T^{-1/2}\varepsilon_t \varepsilon_{t-1}}{\omega_{T\sigma\bar{\gamma}}\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)} \right),$$

where $b_T = \left[T^{\delta}\right]$, $l_T = \left[T^{\theta}\right]$, [$v$] *denotes the largest integer less than v, and* $\theta < \delta <$ 1.

Let $\{\varepsilon_t, \Im_t\}$ *be a strictly stationary martingale difference sequence satisfying the conditions of Theorem 3 except with* $0 < \xi < 1/6$ *and* $\sum_{i=1}^{[T^{1-\delta}]} Y_{Ti}^2 \xrightarrow{p} 1$. *Then*

$$t_{\bar{\gamma}} - \mu(T, c_0, \bar{c}) \xrightarrow{d} N(0,1)$$

where

$$t_{\bar{\gamma}} = \frac{\sum_{t=1}^{T} \Delta y_t \dfrac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}}{\left(\sum_{t=1}^{T} \left(\dfrac{\hat{u}_t \Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2} \right)^2 \right)^{1/2}},$$

$$\mu(T, c_0, \bar{c}) = -\frac{c_0 \omega_{T\gamma\bar{\gamma}}}{\omega_{T\sigma\bar{\gamma}}},$$

and $\hat{u}_t$ *denotes the least squares residuals from the regression in (9).*

PROOF: See Appendix.

The assumptions in Theorems 3 and 4 provide an interesting insight into the trade off between moment and dependence restrictions in this problem. The allowable size of the mixing coefficient is directly related to the rate at which the moments of $\left|\varepsilon_t / \left(cT^{-1/2} + \varepsilon_t^2\right)\right|$ diverge as $T \to \infty$ (see Lemma A1 in the Appendix). This in turn depends on the amount of leptokurtosis in ε_t. If the density function for ε_t is relatively steep as $|\varepsilon_t|$ approaches zero, then $E\left|\varepsilon_t / \left(cT^{-1/2} + \varepsilon_t^2\right)\right|$ diverges faster and the required dependence conditions are tighter.

From Theorem 4, comparing $t_{\bar{\gamma}}$ to a standard normal is the locally best test of a random walk against the alternative $c = \bar{c}$. However, in practice, researchers are unlikely to know the correct value for $\bar{c}$. Further, they will not generally be interested in testing against this specific alternative. They are interested in determining whether there exists a value that provides a significantly better fit than a random walk.

Setting $c = \bar{c}$ and assuming ε_t *iid* normal, we use Theorem 4 to compute the envelope of maximum power. Then, by comparing the power of the test against c_0 for a given $\bar{c}$, we search for values of $\bar{c}$ that yield power close to the envelope. From Figure 2, we see that choosing $\bar{c}$ such that power is optimized somewhere between 50% and 75% causes little loss in overall power. This translates to a choice of $\mu(T,\bar{c},\bar{c})$ between −1.5 and −2.5. Given the Gaussianity assumption, we can integrate to find $\omega_{T\sigma\bar{\gamma}}$, yielding

$$\mu(T,\bar{c},\bar{c}) = -\frac{\bar{c}\,\omega_{T\sigma\bar{\gamma}}}{\sigma_0} = -\gamma^* \sqrt{T}\left(\frac{1+\gamma^*}{\sqrt{\gamma^*}}\sqrt{\pi/2}\,\exp(\gamma^*/2)\left(1-\Phi\left(\sqrt{\gamma^*}\right)\right)-\frac{1}{2}\right)^{\frac{1}{2}}$$

where $\gamma^* = \bar{c}\big/\sqrt{\sigma_0 T}$ and $\Phi(z)$ is the standard normal CDF. Table II lists recommended choices of γ^* for various sample sizes.

Since σ_0^2 is generally unknown, we recommend approximating it with $T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2$. This use of the data in determining γ^* will affect the distribution of the statistic, but we speculate that the effect is small when compared to the benefit of choosing the right order of magnitude for γ^*.

When $\bar{c}$ is set to zero, the $t_{\bar{\gamma}}$ statistic is a variant of the Langrange Multiplier (LM) statistic for a test of H_0: $\gamma_0=0$. From Theorem 4, we see that $\omega_{T\sigma\bar{\gamma}}$ is infinite and $\omega_{T\gamma\bar{\gamma}}$ is finite which implies that $\mu(T,c_0,0)=0$ for all c_0 and therefore the test has no power. The statistic is asymptotically distributed as standard normal under both the null and the alternative. This failure of the LM test results from the null being on the boundary of the parameter space which causes the likelihood function to degenerate. In particular, the derivatives of the likelihood have no finite moments.

All of the above analysis is performed under the assumption of Gaussianity, although it is unlikely that the STOPBREAK disturbances will be Gaussian. Simulating the asymptotic power curves under both a mixture of normals and a GARCH(1,1) reveals a similar picture to that in Figure 2. In fact, the power of the test in these cases is slightly higher due to the excess kurtosis, which helps by driving a wedge between small and large shocks. The size remains correct, as both of these distributions satisfy the assumptions of Theorem 4, given the use of heteroskedasticity consistent standard errors.

4. TESTING IN A MORE GENERAL CONTEXT

In many cases, empirical data will exhibit some temporal correlation in all periods. In the remainder of the paper, we analyze a more general process of the form in (7). This introduces a complexity to the hypothesis testing problem as α_0 is unidentified under the random walk null, yet knowledge of it is required to compute the test statistic. Thus we must choose a value for α_0, despite the fact that it does not exist when the null is true. Suppose we choose the value $\bar{\alpha}$.

We compute the Neyman Pearson test statistic as the t-statistic on ϕ in following regression:

$$(10) \qquad \Delta y_t = \varphi \sum_{i=1}^{t} \bar{\alpha}^{i-1} \frac{\Delta y_{t-i}}{\bar{\gamma} + \Delta y_{t-i}^2} + u_t .$$

As before, the test statistic has an asymptotic normal distribution. This result is given below in Corollary 5.

COROLLARY 5: *Suppose that* $\{y_t\}$, $t = 0, 1, \ldots, T$, *is a stochastic process represented by (7) under the assumptions of Theorem 4 except replacing uniform mixing with strong mixing. Then*

$$t_{\bar{\gamma}} - \mu(T, c_0, \bar{c}, \alpha_0, \bar{\alpha}) \xrightarrow{d} N(0,1)$$

where

$$t_{\bar{\gamma}} = \frac{\sum_{t=1}^{T} \Delta y_t \sum_{i=1}^{t} \bar{\alpha}^{i-1} \frac{\Delta y_{t-i}}{\bar{c}/\sqrt{T} + \Delta y_{t-i}^2}}{\left(\sum_{t=1}^{T} \left(\hat{u}_t \sum_{i=1}^{t} \bar{\alpha}^{i-1} \frac{\Delta y_{t-i}}{\bar{c}/\sqrt{T} + \Delta y_{t-i}^2} \right)^2 \right)^{1/2}},$$

$$\mu(T, c_0, \bar{c}, \alpha_0, \bar{\alpha}) = c_0 (\alpha_0 - 1) \frac{\omega_{T\bar{\alpha}\gamma\bar{\gamma}}}{\omega_{T\bar{\alpha}\bar{\gamma}}},$$

$$\omega_{T\bar{\alpha}\bar{\gamma}}^2 = E\left(\varepsilon_t^2 \sum_{i=1}^{t} \bar{\alpha}^{2(i-1)} \frac{\varepsilon_{t-i}^2}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-i}^2\right)^2} \right),$$

$$\omega_{T\bar{\alpha}\gamma\bar{\gamma}} = E\left(\left(\sum_{i=0}^{\infty} \frac{\alpha_0^i \varepsilon_{t-i}}{\left(c_0/\sqrt{T} + \varepsilon_{t-i}^2\right)} \right) \left(\sum_{i=0}^{\infty} \frac{\bar{\alpha}^i \varepsilon_{t-i}}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-i}^2\right)} \right) \right),$$

and $\hat{u}_t$ *denotes least squares residuals from the regression in (10).*

PROOF: See Appendix.

We see that if $\alpha_0 = 1$, the asymptotic power of the test is equal to the size for all c_0. This also occurs when $\bar{\alpha} = 1$, since in this case $\omega_{T\bar{\alpha}\bar{\gamma}}^2$ is infinite. This lack of power arises because when $\alpha_0 = 1$, the null of a random walk is true. As α_0 goes to one, the correlation within the flat spots increases until they are no longer flat, i.e. the rate at which shocks are expected to decay goes to zero. In contrast, when gamma goes to zero the plateaus shrink in size, keeping the correlation within them at a given level, i.e. we decrease the probability that an observation will have a low permanent impact. Thus, there are two distinct ways of parametrically representing the random walk null.

In order to maximize power, we should choose $\overline{\alpha}$ to be as close as possible to α_0 as often as possible.[6] One strategy could be to choose it arbitrarily. This avoids using the data and thus distorting the distribution of the test statistic. Since the power of the test decreases as α goes to one, it would be advisable to weight this choice of $\overline{\alpha}$ towards one. However if we choose it too close to one, the distribution of the test statistic will be a function of Brownian motions and no longer normal. This arises because the regressor in the test equation becomes nearly integrated as $\overline{\alpha}$ approaches one. A possible choice is $\overline{\alpha} = 0.8$.

A second strategy is to take the infimum of $t_{\overline{\gamma}}$ over feasible values of $\overline{\alpha}$ as suggested by Davies (1977) and elaborated on by Hansen (1996), Andrews (1993) and others. The null distribution of this statistic is well defined but will depend on the correlation of $t_{\overline{\gamma}}$ across various different values of $\overline{\alpha}$ which will in general depend on the distribution of ε_t. In this case, the simulated values prove reasonably robust to GARCH(1,1) and excess kurtosis. We conduct the test using $\overline{\alpha} \in [0, 0.9]$.

Finally, we could approximate the test by regressing Δy_t on $\Delta y_{t-i} / \left(\overline{\gamma} + \Delta y_{t-i}^2\right)$, $i=1,2,\ldots,p$, where p is some predetermined number. Under the null hypothesis TR^2 from this regression will be distributed as $\chi^2_{(p)}$. This procedure will lose power because of the approximation and also because it is unable to test against a one sided alternative, but it is relatively simple to perform.

5. QUASI MAXIMUM LIKELIHOOD ESTIMATION

Given invertibility, we can specify the Gaussian conditional log likelihood function for the process in (7) as

(11) $$L\left(y^T,\varphi\right)=-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\Delta y_t-\alpha\Delta y_{t-1}+\theta_{t-1}\varepsilon_{t-1}\right)^2-\frac{T}{2}\log\left(2\pi\sigma^2\right),$$

where $\varphi=(\gamma\ \ \alpha\ \ \sigma)'$ and $y^T = (y_1, y_2, \ldots, y_T)$. The scores are given by

(12) $$\frac{\partial L}{\partial\gamma}=-\frac{1}{\sigma^2}\sum_{t=1}^{T}\varepsilon_t w_t,$$

(13) $$\frac{\partial L}{\partial\alpha}=-\frac{1}{\sigma^2}\sum_{t=1}^{T}\varepsilon_t v_t,$$

(14) $$\frac{\partial L}{\partial\sigma}=\frac{1}{\sigma^3}\sum_{t=1}^{T}\varepsilon_t^2-T\sigma^{-1},$$

where

(15) $$w_t=b_{t-1}w_{t-1}-(1-\alpha)\frac{\partial q_{t-1}}{\partial\gamma}\varepsilon_{t-1},$$

(16) $$v_t=b_{t-1}v_{t-1}+q_{t-1}\varepsilon_{t-1}-\Delta y_{t-1},$$

(17) $$b_{t-1}=1-(1-\alpha)\left(1+\eta_{q\,t-1}\right)q_{t-1},$$

and $w_0 = v_0 = 0$ wp 1.

Though the likelihood is constructed to be Gaussian, it is not necessary that this be the true distribution. In fact, the presence of large permanent shocks would indicate that leptokurtic errors are likely. We derive consistency and asymptotic normality results below under minimal assumptions on the errors.

THEOREM 6: *Suppose that* $\{y_t\}$ *is a stochastic process represented by (7) with* $\{\varepsilon_t, \Im_t\}$ *a strictly stationary ergodic martingale difference sequence with finite variance* $\sigma_0^2>0$. *Assume that* $q_t(\gamma)$ *is a general function, bounded by zero and one, such that* $E(q_t\varepsilon_t|\Im_{t-1})=0$ *and that the conditions of Theorem 1 are satisfied. Let* Ψ *be a compact subset of* $(-\infty,\infty)\times[0,1)\times(0,\infty)$ *and define* $\hat{\varphi}' = \arg\max_{\Psi} L(y^T,\varphi)$ *and* $\varphi_0 = \arg\max_{\Psi} E(L(y^T,\varphi))$, *where* $\varphi \in \Psi$ *and* φ_0 *is unique. Then* $\hat{\varphi} - \varphi_0 \xrightarrow{\text{a.s.}} 0$.

PROOF: See Appendix.

Given the regularity conditions of Theorem 6, asymptotic normality of QMLE requires further restrictions on the memory of the process. We present this result for the parameterization for q_t given in (5).

THEOREM 7: *Given the conditions of Theorem 6,* $E(\varepsilon_0^2|\Im_{-m}) \xrightarrow{q.m.} \sigma_0^2$ *as* $m\to\infty$, *and* $\sum_{j=0}^{\infty}(\operatorname{var}\Re_{0j})^{1/2}<\infty$ *where* $\Re_{0j} \equiv E(\varepsilon_0^2|\Im_{-j}) - E(\varepsilon_0^2|\Im_{-j-1})$. *Then*

$$V^{-1/2}HT^{1/2}(\hat{\varphi}-\varphi_0) \xrightarrow{d} N(0,I),$$

where $V=\lim_{T\to\infty}\operatorname{cov}(T^{-1/2}\nabla_{\varphi}L(y^T,\varphi_0))$, $\nabla_{\varphi}L(y^T,\varphi)$ *represents the vector of first derivatives of* $L(y^T,\varphi)$, *and*

$$H = \begin{bmatrix} H_1 & 0 \\ 0' & H_2 \end{bmatrix}$$

with $H_1 = -T^{-1}\sigma_0^{-2}\sum_{t=1}^{T} E(s_t s_t')$, $H_2 = 2/\sigma_0^2$ *, and* $s_t = (w_t, v_t)'$. *Assume that H and V are nonsingular and positive definite.*

PROOF: See Appendix.

The result in Theorem 7 may be of limited use in finite samples, especially if the parameters are close to their bounds. For example, the gradient of the likelihood function approaches infinity as γ goes to zero causing confidence intervals to become condensed on the lower side. The linear approximation that yields the asymptotic normality result may thus give misleading confidence regions. A similar scenario arises when α is close to one.

Alternatively, confidence intervals can be formed by inverting a likelihood ratio statistic as in Schoenberg (1997) and Cook and Weisburg (1990). This approach requires finding the value of the parameter such that the likelihood is significantly different from the unconstrained likelihood, where significance is determined by the usual chi-square distribution. Since the likelihood is explicitly used, this method better accommodates the nonlinearity and asymmetry in the model.

The consistency and asymptotic normality results presented above do not rely on the function q_t being correctly specified. From Assumption 2.3 of White (1994), all that is required is that the model density is measurable with respect to the data generating density. Thus if the chosen specification incorrectly represents the true function $q_t = q(\varepsilon_t)$, QMLE estimates remain consistent and asymptotically normal for the KLIC[7] optimal parameter values.

6. APPLICATIONS

6.1. *Two Shock Process*

Consider the process in (3), i.e. $y_t = m_t + \varepsilon_t$ where $m_t = m_{t-1} + u_t$, $prob(u_t=0) = (1-p_u)$, $prob(u_t \sim N(0, \sigma_u^2)) = p_u$, and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Suppose $\sigma_\varepsilon^2 = 1$. Using simulated data and a variety of values for the parameters p_u and σ_u^2, we compare the mean square forecast errors for a number of potential modeling approaches.

Forming conditional forecasts for this process is difficult because it is not invertible. The parameters can be trivially estimated via method of moments, which enables unconditional inference. Shephard (1994) and others have proposed computationally intensive simulation techniques for computing the maximum likelihood estimates. Though a STOPBREAK model is only an approximation, we contend that it is very useful for producing conditional forecasts.

We consider three models; a random walk, an exponential smoother and STOPBREAK. Each is compared to the result from an omniscient modeler who is able to recognize both when a permanent shock has occurred and its magnitude.

The experiment is conducted over 100 trials of 6000 observations each. The STOPBREAK model and the exponential smoother are estimated over the first 5000 observations. One step ahead forecasts are then computed for the next 1000 observations without re-estimation of the model parameters. Mean square forecast errors (MSFE) are computed and compared with analytically calculated MSFEs from a random walk forecast. The results are shown in Table III.

The far right column in Table III contains the difference between the average MSFEs for the exponential smoother and STOPBREAK, with associated standard errors. We see that the relative performance of the STOPBREAK improves with the variance of the permanent shocks. When the variance of the permanent shocks is high, the exponential smoother is penalized for it's slow adjustment. In other words, this scenario highlights the ability of the STOPBREAK model to react quickly to large permanent innovations. As the permanent shocks become smaller on average, the performance of the two models becomes insignificantly different. These results are largely independent of the frequency of the breaks though in some cases, the superiority of the STOPBREAK model is less significant at low break frequencies. This likely arises because the more permanent breaks there are, the more opportunity the STOPBREAK model has to exercise its comparative advantage.

6.2. *Relative Stock Prices*

Individual stock prices have a tendency to move together, by virtue of their existence in a common market. This empirical observation is backed up by a number of asset pricing models, the most well known being the CAPM where individual stock returns are proportional to the return on the market portfolio. However these models tell us little about the dynamic features of relative stock prices. The goal of this analysis is to gain insight into the dynamics of relations between stock prices over time.

Many researchers have conducted cointegration tests amongst asset prices, with mixed results. Chelley-Steeley and Pentecost (1994) and Cerchi and Havenner (1988) both find evidence of cointegration among stock prices, while Stengos and Panas (1992) and MacDonald and Power (1991) do not. Granger (1986) documents that the predictability inherent in cointegrating relationships precludes them from existing in an efficient market.

However it may be that such relationships could exist over short periods of time in a market with imperfect information. Consider a market in which two stocks, A and B, are traded. Suppose that firms A and B are in the same industry. Various types of information will lead to a change in the value of just one stock or of both. Whenever the price of one stock changes, investors must ascertain whether this information is relevant in determining the value of the other stock. For example, if firm A announces better than expected profits, it may indicate higher profits for the whole industry. The price of stock B will initially rise to reflect this possibility and then as more information becomes available, it may go to a new 'high profit' level, in which case the ratio returns to its original value. Otherwise stock B may return to its initial point leaving the ratio permanently at a new level. Thus we observe serial correlation in the relative price of the stocks with a time varying permanent impact of shocks.

As an example, the logarithm of the Mobil to Texaco relative share price is shown in Figure 3. The presence of apparently stationary sections indicates periods where shocks have a low permanent impact punctuated by episodes of a high impact. This is most evident in the last five years of the sample where there are a number of three to six month periods where the first order autocorrelation in the series is around 0.7.

We apply the STOPBREAK model to daily data on a number of stocks over a sample period of January 1988 through December 1995. The data were obtained from Datastream International.

The results of tests of the random walk null are presented in Table IV, where each column in the table contains the statistic computed using different methods of accommodating the unidentified parameter, as outlined in Section 5. We see that the null is rejected in almost all cases, with the conclusion generally invariant to the method of setting the unidentified parameter.

QMLE estimates of the model over the full sample are given in Table V for four stock pairs. The estimates of α and the standardized γ are broadly similar across the four pairs, although confidence bands are wider for Coke/Pepsico and J&J/Merck. In both of these cases, there is a high positive correlation between $\hat{\alpha}$ and $\hat{\gamma}$. Intuitively, the model is having difficulty distinguishing between high temporal correlation with few breaks and lower correlation with more breaks.

In all but one case the estimates of γ are within two standard deviations of zero, when heteroskedasticity consistent standard errors are used. However 95% confidence intervals computed through inverting the appropriate likelihood ratio statistic reveal a much shorter interval on the lower end.

Comparing the estimated variance of the STOPBREAK errors with the estimated variance of Δy_t reveals a low fit, with R^2 often less than 1%. The low fit is not unexpected, given the nature of financial data. However, it may possibly be improved by a richer specification of the model. For example, we could estimate some equilibrium relationship between the stock prices rather than using the ratio. Incorporating features such as conditional heteroskedasticity and more memory in q_t could also prove fruitful.

A further indicator of the presence of stochastic permanent breaks in stock price ratios is to compute potential profits from a trading strategy. The STOPBREAK model predicts the direction that the price ratio will move, i.e. it forecasts whether the prices will move towards or away from each other. If they are predicted to move apart, the investor will buy the higher valued stock and sell the lower valued stock short. In a STOPBREAK framework, such an investor is expected to make small gains regularly and then to make either large gains or large losses when the unexpected permanent shocks occur. On average, these large profits and losses will cancel each other out, leaving an accumulated wealth with no money down.

We compute the profits gained from enacting the above pairs trading strategy for January 1996 through June 1997. The model is not re-estimated during the forecast period and standard errors are computed assuming that daily profits are *iid*. The results are given in Table VI.

As a benchmark, we compare the STOPBREAK profits with those from using the same strategy, but forecasting using an exponential smoother and a 20 day moving average. For J&J/Merck and Mobil/Texaco, STOPBREAK outperforms the other models both in terms of mean return and the Sharpe Ratio[8]. For the other two cases, the naïve moving average model performs better. In no cases does the exponential smoother beat the STOPBREAK model.[9]

7. CONCLUSION

We have proposed a new approach to modeling processes where the effect of shocks fluctuates between permanent and transient. Typically, such data exhibits periods of apparent stationarity punctuated by occasional permanent mean shifts. Rather than attempting to individually characterize a number of different regimes, we allow the process to predict whether part or all of a shock will be permanent or transitory. The stochastic permanent breaks (STOPBREAK) process assumes that permanent shocks can be identified by their larger magnitude, but this assumption should be viewed as a special case rather than a necessity.

We considered two applications of a STOPBREAK model. Firstly, we analyzed simulated *iid* data with random mean shifts of random amounts. The STOPBREAK model can be viewed as an approximation to this type of process. Out of sample mean square forecast errors reveal STOPBREAK to be on a par with the exponential smoother when the variance of the permanent shifts is small. As the size of the permanent shock increases, the relative performance of STOPBREAK improves and it becomes significantly better.

The second application involved pairs of stock prices. We posit that the relative price of two stocks follows a STOPBREAK process. The two prices tend to move together for periods of time and jump apart occasionally. Hypothesis tests reveal evidence of stochastic permanent breaks and maximum likelihood estimates are used to form a profitable out of sample trading strategy.

Department of Economics, University of California, San Diego, La Jolla, CA 92093, U.S.A.

APPENDIX

PROOF OF THEOREM 1: Let $\{\tilde{y}_t\}$, t = 0, 1,…, T, be a sequence of realizations from the process in (1) and (2) and write

(A.1) $$\tilde{\varepsilon}_t = (1-\tilde{q}_{t-1})\tilde{\varepsilon}_{t-1} + \Delta\tilde{y}_t \qquad (t = 1, 2,\ldots, T).$$

where $\{\tilde{\varepsilon}_t\}$ denotes the sequence of realizations from $\{\varepsilon_t, \Im_t\}$ that generated $\{\tilde{y}_t\}$.

Invertibility requires that the innovations can be computed uniquely from the observed time series. Consider the sequence of real numbers, $\hat{\varepsilon}_0, \hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T$, where

(A.2) $$\hat{\varepsilon}_t = (1-\hat{q}_{t-1})\hat{\varepsilon}_{t-1} + \Delta\tilde{y}_t \qquad (t = 1, 2,\ldots, T).$$

and $\hat{q}_{t-1} = q(\hat{\varepsilon}_{t-1})$. Suppose that $\hat{\varepsilon}_0$ is drawn from some unbounded continuous distribution.

From Granger and Andersen (1978), the process is invertible if the sequence $\{\tilde{\varepsilon}_t - \hat{\varepsilon}_t\}$ converges to zero. Thus, since the recursions generating $\{\tilde{\varepsilon}_t\}$ and $\{\hat{\varepsilon}_t\}$ are identical, it is sufficient to show that the starting value in (A.2) has no effect in the limit. The result here could be termed “weak invertibility”, as we show only that convergence occurs wp 1.

Consider the effect on $\hat{\varepsilon}_1$ of a change in $\hat{\varepsilon}_0$:

$$\frac{\partial\hat{\varepsilon}_1}{\partial\hat{\varepsilon}_0} = 1 - \left(\hat{q}_0 + \left.\frac{\partial q}{\partial\hat{\varepsilon}}\right|_{\hat{\varepsilon}_0} \hat{\varepsilon}_0\right).$$

This leads to

$$\frac{\partial\hat{\varepsilon}_T}{\partial\hat{\varepsilon}_0} = \prod_{t=1}^{T}\frac{\partial\hat{\varepsilon}_t}{\partial\hat{\varepsilon}_{t-1}} = \prod_{t=0}^{T-1}\left(1 - \left(\hat{q}_t + \left.\frac{\partial q}{\partial\hat{\varepsilon}}\right|_{\hat{\varepsilon}_t} \hat{\varepsilon}_t\right)\right)$$

$$= \prod_{t=0}^{T-1}\left(1-\hat{q}_t\left(1+\hat{\eta}_{qt}\right)\right).$$

Since $0 \le q_t \le 1$ $\forall t$ and $\hat{q}_t$ is non-decreasing in $|\hat{\varepsilon}_t|$,

$$\left.\frac{\partial q}{\partial\varepsilon}\right|_{\hat{\varepsilon}_t} \hat{\varepsilon}_t \ge 0 \ \text{ wp 1 } \forall t,$$

which implies $\hat{\eta}_{qt} \ge 0$ $\forall t$. Further, given that $q_t(1+\eta_{qt}) < 2$ wp 1 $\forall t$, and since $\hat{\varepsilon}_0$ is drawn from some continuous distribution, sequences $\{\hat{\varepsilon}_t\}$ such that $\hat{q}_t(1+\hat{\eta}_{qt}) \ge 2$ for some t occur on a set of measure zero. This yields

$$-1 < \frac{\partial \hat{\varepsilon}_T}{\partial \hat{\varepsilon}_0} \le 1 \quad \text{wp 1.}$$

It remains to show that $\hat{q}_t(1+\hat{\eta}_{qt})$ is greater than zero with positive probability. Since

$$\left.\frac{\partial q}{\partial \varepsilon}\right|_{\varepsilon_t} \varepsilon_t \ge 0 \quad \text{wp 1 } \forall t,$$

and $q_t > 0$ with positive probability, there exists $k < \infty$ such that $q(\varepsilon_t) > 0$ for all $|\varepsilon_t| > k$. Thus, the process is invertible if prob($|\hat{\varepsilon}_t| > k$) > 0. Suppose $|\hat{\varepsilon}_t| \le k$ wp 1 $\forall t$. This implies

$$\hat{\varepsilon}_T = \sum_{t=1}^{T} \Delta \tilde{y}_t = \tilde{y}_T - \tilde{y}_0,$$

but

$$\lim_{T\to\infty} \text{prob}\left(|y_T - y_0| < k\right) = 0$$

since, using the martingale difference sequence property of $\{\varepsilon_t, \Im_t\}$ and $\{q_t\varepsilon_t, \Im_t\}$,

$$\lim_{T\to\infty} \text{var}(y_T - y_0) = \lim_{T\to\infty} E\left(\sum_{t=1}^{T} q_{t-i}\varepsilon_{t-i} + \varepsilon_T - \varepsilon_0\right)^2$$

$$= \lim_{T\to\infty}\left(\sum_{t=1}^{T} E(q_{t-i}\varepsilon_{t-i})^2 + E(\varepsilon_T^2) - E(\varepsilon_0^2)\right) = \infty .$$

Thus the set of sequences $\{\hat{\varepsilon}_t\}$ such that $|\hat{\varepsilon}_t| \le k$ $\forall t$ occurs on a set of measure zero.

This implies

$$\lim_{T\to\infty}\left(\text{prob}\left(\left|\frac{\partial \hat{\varepsilon}_T}{\partial \hat{\varepsilon}_0}\right| < \delta\right)\right) = 1$$

for every $\delta>0$, and the process is invertible wp 1. *Q.E.D.*

PROOF OF THEOREM 2: Write the STOPBREAK process in (1) and (2) as

$$y_t = -\gamma_0 \sum_{i=1}^{t} u_{i-1} + \sum_{i=1}^{t} \varepsilon_i + m_0$$

where $u_t = \varepsilon_t / (\gamma_0 + \varepsilon_t^2)$ and m_0 is fixed and known. Since

$$|u_t| \le \frac{1}{2\sqrt{\gamma_0}} \ \ \forall t,$$

all moments of u_t exist. Symmetry of the distribution of $\varepsilon_t|\Im_{t-1}$ about zero implies that $u_t|\Im_{t-1}$ is also distributed symmetrically about zero. Thus $\{u_t, \Im_t\}$ is a martingale difference sequence.

We use the functional central limit theorem in Phillips and Solo (1992, Theorem 2.6). This requires verifying that $T^{-1}\sigma_0^{-2}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} 1$ and $\max_{1\le t\le T} T^{-1/2}\sigma_0^{-1}|\varepsilon_t| \xrightarrow{p} 0$, and similarly for u_t.

Strict stationarity and ergodicity of ε_t implies that ε_t^2 and u_t are also stationary and ergodic (see White (1984, Theorem 3.35)). Then, from the ergodic theorem (White (1984, Theorem 3.34)), and the result that almost sure convergence implies convergence in probability (White (1984, Theorem 2.24)), we have $T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} \sigma_0^2$ and $T^{-1}\sum_{t=1}^{T}u_t^2 \xrightarrow{p} \sigma_0^2$.

For any δ>0,

$$\Pr\mathrm{ob}\left(\max_{1\le t\le T} T^{-1/2}\sigma_0^{-1}|\varepsilon_t| < \delta\right) = \Pr\mathrm{ob}\left(T^{-1/2}\sigma_0^{-1}\max_{1\le t\le T}|\varepsilon_t| < \delta\right)$$

$$\ge \Pr\mathrm{ob}\left(T^{-1/2}\sigma_0^{-1}|\varepsilon_t| < \delta\right)$$

$$\ge 1 - \frac{T^{-1/2}\sigma_0^{-1}E|\varepsilon_t|}{\delta}$$

from the generalized Chebyshev inequality. Thus

$$\Pr\mathrm{ob}\left(\max_{1\le t\le T} T^{-1/2}\sigma_0^{-1}|\varepsilon_t| < \delta\right) \to 1 \text{ as } T \to \infty.$$

Boundedness of u_t ensures that $\max_{1\le t\le T} T^{-1/2}\sigma_0^{-1}|u_t| \xrightarrow{p} 0$.

Then invoking the functional central limit theorem yields

$$T^{-1/2}(y_t - m_0) \xrightarrow{d} -\gamma_0\omega_0 W_\gamma(r) + \sigma_0 W(r),$$

where $t=[rT]$, $[rT]$ denotes the largest integer below rT and $W(r)$ and $W_\gamma(r)$ are dependent non-identical standard Brownian motions on the unit interval. Henceforth, we define $W \equiv W(r)$ and $W_\gamma \equiv W_\gamma(r)$ for brevity.

We have

$$T\hat{\rho} = \frac{T^{-1}\sum_{t=1}^{T} \Delta y_t y_{t-1}}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}.$$

The numerator is

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T} y_{t-1}\Delta y_t &= T^{-1}\sum_{t=2}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i - \gamma_0\sum_{i=1}^{t-1}\frac{\varepsilon_{i-1}}{\gamma_0+\varepsilon_{i-1}^2} + m_0\right)\left(\varepsilon_t - \frac{\gamma_0\varepsilon_{t-1}}{\gamma_0+\varepsilon_{t-1}^2}\right) + T^{-1}(m_0+\varepsilon_0)\left(\varepsilon_1 - \frac{\gamma_0\varepsilon_0}{\gamma_0+\varepsilon_0^2}\right) \\
&= T^{-1}\sum_{t=2}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i - \gamma_0\sum_{i=1}^{t-1}\frac{\varepsilon_{i-1}}{\gamma_0+\varepsilon_{i-1}^2}\right)\left(\varepsilon_t - \frac{\gamma_0\varepsilon_{t-1}}{\gamma_0+\varepsilon_{t-1}^2}\right) + o_p(1) \\
&= T^{-1}\sum_{t=2}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i - \gamma_0\sum_{i=1}^{t-1}\frac{\varepsilon_{i-1}}{\gamma_0+\varepsilon_{i-1}^2}\right)\varepsilon_t \\
&\qquad - \gamma_0 T^{-1}\sum_{t=2}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i - \gamma_0\sum_{i=1}^{t-1}\frac{\varepsilon_{i-1}}{\gamma_0+\varepsilon_{i-1}^2}\right)\frac{\varepsilon_{t-1}}{\gamma_0+\varepsilon_{t-1}^2} + o_p(1) \\
&\xrightarrow{d} \int(\sigma_0 W - \gamma_0\omega_0 W_\gamma)(\sigma_0\, dW) - \gamma_0\int(\sigma_0 W - \gamma_0\omega_0 W_\gamma)(\omega_0\, dW_\gamma)
\end{aligned}$$

from the continuous mapping theorem (see, for example, Davidson (1994, Theorem 26.13)). For the denominator, we have

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 = T^{-1}\sum_{t=1}^{T}\left(T^{-1/2}y_{t-1}\right)^2 \xrightarrow{d} \int(\sigma_0 W - \gamma_0\omega_0 W_\gamma)^2$$

from the continuous mapping theorem and the result follows. *Q.E.D.*

The following Lemma will be useful in proving Theorems 3, 4 and 5.

LEMMA A1: *Let X_t be some strictly stationary positively valued random variable defined on a complete probability space. Consider the function*

$$f_{Tt} = \frac{X_t}{c/\sqrt{T} + X_t^2}$$

where $0 < c < \infty$*. Then, for any* $\eta > 0$, $E\left(f_{Tt}^{1+\eta}\right) = O\left(T^{\xi(1+\eta)}\right)$ *where* $0 < \xi < 1/4$.

PROOF OF LEMMA A1: For $T < \infty$, we have

$$\frac{X_t}{c/\sqrt{T} + X_t^2} \le \frac{T^{1/4}}{2c^{1/2}}.$$

Let $I(A>B)$ be a function taking the value one if $A > B$ and zero otherwise. Now, for some $0 < \delta < 1/4$, consider

$$E\left(f_{Tt}^{1+\eta}\right) = E\left(f_{Tt}^{1+\eta} I\left(f_{Tt} \leq T^{1/4-\delta}/2c^{1/2}\right)\right) + E\left(f_{Tt}^{1+\eta} I\left(f_{Tt} > T^{1/4-\delta}/2c^{1/2}\right)\right)$$

$$\leq \left(\frac{T^{1/4-\delta}}{2c^{1/2}}\right)^{1+\eta} + E\left(f_{Tt}^{1+\eta} I\left(f_{Tt} > T^{1/4-\delta}/2c^{1/2}\right)\right)$$

$$\leq \left(\frac{T^{1/4-\delta}}{2c^{1/2}}\right)^{1+\eta} + \left(\frac{T^{1/4}}{2c^{1/2}}\right)^{1+\eta} \mathrm{Prob}\left(f_{Tt} > T^{1/4-\delta}/2c^{1/2}\right)$$

$$= \left(\frac{T^{1/4-\delta}}{2c^{1/2}}\right)^{1+\eta} + \left(\frac{T^{1/4}}{2c^{1/2}}\right)^{1+\eta} \mathrm{Prob}\left(X_t \in [A_T, B_T]\right)$$

where

$$A_T = \left(c^{1/2}/T^{1/4-\delta}\right)\left(1-\left(1-T^{-2\delta}\right)^{1/2}\right) = O\left(T^{\delta-1/4}\right), \text{ and}$$

$$B_T = \left(c^{1/2}/T^{1/4-\delta}\right)\left(1+\left(1-T^{-2\delta}\right)^{1/2}\right) = O\left(T^{\delta-1/4}\right).$$

Now since $\mathrm{Prob}\left(X_t \in [A_T, B_T]\right) \leq 1$ there exists $\theta > 0$ such that $\mathrm{Prob}\left(X_t \in [A_T, B_T]\right) = O\left(T^{(\delta-1/4)\theta}\right)$. Thus we have

$$E\left(f_{Tt}^{1+\eta}\right) \leq O\left(T^{(1/4-\delta)(1+\eta)}\right) + O\left(T^{1/4(1+\eta)+(\delta-1/4)\theta}\right) = O\left(T^{\xi(1+\eta)}\right)$$

where

$$\xi = \max\left(1/4-\delta,\ 1/4 - \frac{(\delta-1/4)\theta}{1+\eta}\right) < 1/4 . \qquad \textit{Q.E.D.}$$

PROOF OF THEOREM 3: Define

$$w_{Tt} = \frac{\varepsilon_t^2}{\left(c/\sqrt{T} + \varepsilon_t^2\right)^2} \qquad (t = 0, 1,\ldots, T),$$

where $0 < c < \infty$. We show that $\left\{\left(w_{Tt} - E(w_{Tt})\right), \Im_{Tt}\right\}$ is an L_1-mixingale, and use the law of large numbers in Theorem 1 of de Jong (1995). Since w_{Tt} is measurable-$\Im_{Tt}$, $\left\{\left(w_{Tt} - E(w_{Tt})\right), \Im_{Tt}\right\}$ is an L_1-mixingale if there exists an array of non-negative constants $\{s_{Tt}\}_{-\infty}^{\infty}$ and a non-negative sequence $\{\zeta_j\}_0^{\infty}$ such that $\zeta_j \to 0$ as $j \to \infty$ and

$$E\left|E\left(w_{Tt}-E(w_{Tt})\middle|\Im_{t-j}\right)\right|\le s_{Tt}\zeta_j$$

(see, for example, Davidson (1994, Definition 16.5)). From Davidson (1994, Theorem 14.4),

$$E\left|E\left(w_{Tt}\middle|\Im_{t-j}\right)-E(w_{Tt})\right|\le 2\left(E\left(w_{Tt}^{1+\eta}\right)\right)^{1/(1+\eta)}\phi_w(j)^{\eta/(1+\eta)}$$

for some $\eta > 0$, where $\phi_w(j)$ denotes the uniform mixing coefficient for w_{Tt}. Define

$$s_{Tt}\equiv 2\left(E\left(w_{Tt}^{1+\eta}\right)\right)^{1/(1+\eta)} \text{ and } \zeta_j\equiv\phi_w(j)^{\eta/(1+\eta)}.$$

Now from Lemma A1,

$$\left(E\left(w_{Tt}^{1+\eta}\right)\right)^{1/(1+\eta)}=O\left(T^{2\xi}\right)$$

where $\xi < 1/4$.

Since w_{Tt} is an $\Im_t$ -measurable function of finite length, we have from White (1984, Theorem 3.49) that $\{w_{Tt},\Im_{Tt}\}$ is mixing of the same size as $\{\varepsilon_t,\Im_t\}$. Then, given finite η, $\zeta_j\to 0$ as $j\to\infty$ and therefore $\{(w_{Tt}-E(w_{Tt})),\Im_{Tt}\}$ is an L_1-mixingale.

Given stationarity of $\{\varepsilon_t,\Im_t\}$, the mixingale law of large numbers requires that, for some sequence $C_T\geq 1$, $C_T=o(T^{1/2})$,

(A) $\quad \lim\limits_{B\to\infty}\limsup\limits_{T\to\infty}E\left|X_{Tt}\ I\left(\left|X_{Tt}\right|>BC_T\right)\right|=0$ and,

(B) $\quad$ for all $K>0$, $\lim\limits_{T\to\infty}s_{Tt}\zeta\left(\left[KT^{1/2}C_T^{-1}\right]\right)=0$,

where $I\left(\left|X_{Tt}\right|>BC_T\right)$ takes the value one if $\left|X_{Tt}\right|>BC_T$ and zero otherwise. In this case, we have $X_{Tt}=w_{Tt}-E(w_{Tt})$.

Consider $C_T = T^{2\xi}$. Now

$$E\left|X_{Tt}\ I\left(\left|X_{Tt}\right|>BC_T\right)\right|=E\left|\left(w_{Tt}-E(w_{Tt})\right)\ I\left(\left|X_{Tt}\right|>BT^{2\xi}\right)\right|$$

$$\le 2E\left(w_{Tt}\ I\left(w_{Tt}>BT^{2\xi}/2\right)\right)+2\left|E(w_{Tt})I\left(E(w_{Tt})>BT^{2\xi}/2\right)\right|$$

from Davidson (1994, Theorem 9.29) and since $\left|w_{Tt}-E(w_{Tt})\right|\le w_{Tt}+E(w_{Tt})$ by the triangle inequality. Since

$$\left(E\left(w_{Tt}^{1+\eta}\right)\right)^{1/(1+\eta)}=O\left(T^{2\xi}\right) \text{ and } E(w_{Tt})\le\left(E\left(w_{Tt}^{1+\eta}\right)\right)^{1/(1+\eta)},$$

we have $E(w_{Tt})=O(T^{2\xi})$ and therefore there exist B and N such that $E(w_{Tt})\le BT^{2\xi}$ for all $T \geq N$, i.e.

$$\lim_{B\to\infty}\limsup_{T\to\infty}\left(E(w_{Tt})I\left(E(w_{Tt})>BT^{2\xi}\right)\right)=0\,.$$

Consider

$$E(w_{Tt})=E\left(w_{Tt}I\left(w_{Tt}\le BT^{2\xi}\right)\right)+E\left(w_{Tt}I\left(w_{Tt}>BT^{2\xi}\right)\right)$$

$$\le BT^{2\xi}+E\left(w_{Tt}I\left(w_{Tt}>BT^{2\xi}\right)\right)$$

Now since $E(w_{Tt})=O(T^{2\xi})$, we have

$$\lim_{B\to\infty}\limsup_{T\to\infty}E\left(w_{Tt}I\left(w_{Tt}>BT^{2\xi}\right)\right)=0\,.$$

Thus condition (A) is satisfied.

For the second condition, we have

$$\lim_{T\to\infty}s_{Tt}\zeta\left(\left[KT^{1/2}C_T^{-1}\right]\right)=\lim_{T\to\infty}T^{-2\xi}s_{Tt}T^{2\xi}\zeta\left(\left[KT^{1/2-2\xi}\right]\right).$$

Let $j = T^{1/2-2\xi}$, which implies that $T^{2\xi} = j^{\xi/(1/4-\xi)}$. Thus

$$\lim_{T\to\infty}T^{2\xi}\phi\left(\left[KT^{1/2-2\xi}\right]\right)^{\eta/(1+\eta)}=\lim_{j\to\infty}j^{\frac{\xi}{1/4-\xi}}\phi\left(\left[Kj\right]\right)^{\eta/(1+\eta)}=0$$

since $\phi(j) = o(j^{-\kappa})$ where

$$\kappa=\left(\frac{1+\eta}{\eta}\right)\left(\frac{\xi}{1/4-\xi}\right).$$

Now, since $s_{Tt} = O(T^{2\xi})$, we have

$$\lim_{T\to\infty}s_{Tt}\zeta\left(\left[KT^{1/2}C_T^{-1}\right]\right)=0\,,$$

and can apply the law of large numbers to obtain the result. *Q.E.D.*

PROOF OF THEOREM 4: We have

$$t_{\bar{\gamma}}=\frac{\dfrac{1}{\sqrt{T}}\displaystyle\sum_{t=1}^{T}\Delta y_t\,\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}}{\left(\dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\left(\frac{\hat{u}_t\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)^2\right)^{1/2}}\,.$$

Note that

$$\Delta y_{t-1} = \varepsilon_{t-1} - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T} + \varepsilon_{t-2}^2}\varepsilon_{t-2}$$

and consider the numerator of $t_{\bar{\gamma}}$.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Delta y_t \frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \frac{\left(\varepsilon_t - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T} + \varepsilon_{t-1}^2}\varepsilon_{t-1}\right)\left(\varepsilon_{t-1} - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T} + \varepsilon_{t-2}^2}\varepsilon_{t-2}\right)}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right) - \frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-2}}{\left(c_0/\sqrt{T} + \varepsilon_{t-2}^2\right)\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right)$$

$$- \frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^2}{\left(c_0/\sqrt{T} + \varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right) + \frac{c_0^2}{T^{3/2}}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}\varepsilon_{t-2}}{\left(c_0/\sqrt{T} + \varepsilon_{t-1}^2\right)\left(c_0/\sqrt{T} + \varepsilon_{t-2}^2\right)\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right)$$

We analyze each of these four terms separately, labeling them (i), (ii), (iii), and (iv) respectively.

(i)
$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2}\right) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}\left(\varepsilon_{t-1}^2 - \Delta y_{t-1}^2\right)}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right)$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2}\right) + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}\left(2c_0\varepsilon_{t-1}\varepsilon_{t-2}\right)}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-2}^2\right)\left(\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right)$$

$$- \frac{1}{T^{3/2}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}\left(c_0^2\varepsilon_{t-2}^2\right)}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-2}^2\right)^2\left(\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right)$$

$$\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2}\right) + 2c_0\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t U_{Tt} - c_0^2\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t V_{Tt}$$

where $\{\varepsilon_t U_{Tt}, \Im_{Tt}\}$ and $\{\varepsilon_t V_{Tt}, \Im_{Tt}\}$ are martingale difference arrays. We proceed in a similar fashion to the Proof of Theorem 3. One notable difference relates to terms with Δy_t^2 rather than ε_t^2 in the denominator. Now since

$$\left|\Delta y_{t-1}\right| = \left|\varepsilon_{t-1} - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T} + \varepsilon_{t-2}^2}\varepsilon_{t-2}\right|,$$

$$\geq |\varepsilon_{t-1}| - \frac{c_0^{1/2}}{2T^{1/4}},$$

we have

$$\left|\frac{\varepsilon_{t-1}}{c_0 / \sqrt{T} + \Delta y_{t-1}^2}\right| \leq \left|\frac{\varepsilon_{t-1}}{c_0 / \sqrt{T} + \left(|\varepsilon_{t-1}| - c_0^{1/2} / 2T^{1/4}\right)^2}\right| \leq \frac{T^{1/4}\sqrt{5}}{6c_0^{1/2}}.$$

Thus this term is bounded and the bound and moments diverge at the same rate as the term in Lemma A1.

Now since $\{\varepsilon_t U_{Tt}, \Im_{Tt}\}$ is a martingale difference array, it is also a mixingale with $s_{Tt} = \left(E|U_{Tt}|^{1+\eta}\right)^{1/(1+\eta)}$ and ζ_j=0, for all $j \neq 0$. From Holders inequality, we have

$$E|U_{Tt}|^{1+\eta} \leq \left(E\left|\frac{\varepsilon_{t-1}}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2\right)}\right|^{(1+\eta)p}\right)^{\frac{1}{p}} \left(E\left|\frac{\varepsilon_{t-1}}{\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right|^{(1+\eta)q}\right)^{\frac{1}{q}} \left(E\left|\frac{\varepsilon_{t-2}}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-2}^2\right)}\right|^{(1+\eta)r}\right)^{\frac{1}{r}}$$

where $p > 1$, $q > 1$ and $1/p + 1/q + 1/r = 1$. This leads to

$$\left(E|U_{Tt}|^{1+\eta}\right)^{1/(1+\eta)} = O\left(T^{3\xi}\right)$$

where $\xi < 1/4$, from Lemma A1. However to use the weak law for mixingales as in the proof of Theorem 3 requires $\xi < 1/6$, which is true by assumption. Substituting $\varepsilon_t U_{Tt}$ for w_{Tt} in the proof of Theorem 3 then yields

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t U_{Tt} \xrightarrow{L_1} 0.$$

Similarly for $\varepsilon_t V_{Tt}$, we have

$$E|V_{Tt}|^{1+\eta} \leq \left(E\left|\frac{T^{-1/2}}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2\right)}\right|^{(1+\eta)p}\right)^{\frac{1}{p}} \left(E\left|\frac{\varepsilon_{t-1}}{\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)}\right|^{(1+\eta)q}\right)^{\frac{1}{q}} \left(E\left|\frac{\varepsilon_{t-2}^2}{\left(\bar{c}/\sqrt{T} + \varepsilon_{t-2}^2\right)^2}\right|^{(1+\eta)r}\right)^{\frac{1}{r}}$$

Now $T^{-1/2} / \left(\bar{c}T^{-1/2} + \varepsilon_{t-1}^2\right) \leq 1/\bar{c} = O(1)$, which implies that

$$\left(E|V_{Tt}|^{1+\eta}\right)^{1/(1+\eta)} = O\left(T^{3\xi}\right)$$

where $\xi < 1/6$ as before. Substituting $\varepsilon_t V_{Tt}$ for w_{Tt} in the proof of Theorem 3 then yields

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t V_{Tt} \xrightarrow{L_1} 0.$$

Thus, for term (i), we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2}\right)+o_p(1).$$

We use the central limit theorem given in Theorem 1 of de Jong (1997). Define

$$Z_{Tt}=\frac{T^{-1/2}\varepsilon_t\varepsilon_{t-1}}{\omega_{T\sigma\bar{\gamma}}\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)}$$

where

$$\omega_{T\sigma\bar{\gamma}}=\left(E\left(\frac{\varepsilon_t^2\varepsilon_{t-1}^2}{\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)^2}\right)\right)^{\frac{1}{2}}.$$

Since $\{\varepsilon_t, \Im_t\}$ is a stationary martingale difference sequence, $\{Z_{Tt}, \Im_{Tt}\}$ is also a martingale difference sequence for all $\bar{c}>0$. From Davidson (1994, Theorem 14.4),

$$\left(E\left(Z_{Tt}\left|\Im_{t-j}\right.\right)^2\right)^{1/2}\leq 2\left(E\left(Z_{Tt}\right)^2\right)^{1/2}\phi(j)^{1/2}=2T^{-1/2}\phi(j)^{1/2}.$$

Then, given that $\phi(j)\to 0$ as $j\to\infty$, $\{Z_{Tt}, \Im_{Tt}\}$ is a L_2-mixingale array of size –1/2 with $s_{Tt}=2T^{-1/2}$ and $\zeta_j=\phi(j)^{1/2}$.

Use of de Jongs (1997) central limit theorem requires that three further conditions be satisfied. The first, that $(Z_{Tt}/s_{Tt})^2$ be uniformly integrable is true if $E(Z_{Tt}/s_{Tt})^{2+\theta}<\infty$ (see Davidson (1994, Theorem 12.10)). We have

$$E\left|\frac{Z_{Tt}}{s_{Tt}}\right|^{2+\theta}=\frac{1}{\omega_{T\sigma\bar{\gamma}}^{2+\theta}}E\left|Z_{Tt}\right|^{2+\theta}=O(1)$$

from Lemma A1. Since $s_{Tt}=T^{-1/2}$ for all t, the second condition is satisfied if $s_{Tt}=o\left(b_T^{-1/2}\right)$ and $\sum_{i=1}^{[T^{1-\delta}]}s_{Tt}^2=O\left(b_T^{-1}\right)$, which follows trivially. Finally, the conditional variance condition, that $\sum_{i=1}^{[T^{1-\delta}]}Y_{Ti}^2\xrightarrow{p}1$ is satisified by assumption. Thus, we have

$$\frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-1}}{\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2}\right)\xrightarrow{d}N(0,1).$$

(ii) Consider

$$\frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t\varepsilon_{t-2}}{\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)}\right) \equiv \frac{c_0}{T}\sum_{t=1}^{T}\varepsilon_t U_{Tt} \ .$$

Now from Holders inequality,

$$E\left|U_{Tt}\right|^{1+\eta} \le \left(E\left|\frac{1}{\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)}\right|^{(1+\eta)q}\right)^{\frac{1}{q}}\left(E\left|\frac{\varepsilon_{t-2}}{\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)}\right|^{(1+\eta)r}\right)^{\frac{1}{r}}$$

where $1/q + 1/r = 1$. This leads to

$$\left(E\left|U_{Tt}\right|^{1+\eta}\right)^{1/(1+\eta)} = O\left(T^{3\xi}\right)$$

from Lemma A1. Substituting $\varepsilon_t U_{Tt}$ for w_{Tt} in the proof of Theorem 3 then yields

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t U_{Tt} \xrightarrow{L_1} 0\,.$$

(iii)
$$\frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^2}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)}\right)$$

$$=\frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^2}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)}\right)+\frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^2\left(\varepsilon_{t-1}^2-\Delta y_{t-1}^2\right)}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)}\right)$$

$$=\frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^2}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)}\right)$$

$$+\frac{2c_0^2}{T^{3/2}}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^3\varepsilon_{t-2}}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)}\right)$$

$$-\frac{c_0^3}{T^2}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^2\varepsilon_{t-2}^2}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)^2}\right)$$

$$\equiv c_o\omega_{T\gamma\bar{\gamma}} + T^{\xi-1/2}\frac{2c_0}{T}\sum_{t=1}^{T}U_{Tt} - T^{3\xi-1}\frac{c_0^3}{T}\sum_{t=1}^{T}V_{Tt} + o_p(1)$$

from Theorem 3. Consider, for $\eta > 0$,

$$E\left|U_{Tt}\right|^{1+\eta} = E\left|\frac{T^{-\xi}\varepsilon_{t-1}^3\varepsilon_{t-2}}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)}\right|^{1+\eta}$$

$$= O\left(T^{-\xi(1+\eta)}\right)O\left(T^{4\xi(1+\eta)}\right)$$

We have $E\left|U_{Tt}\right|^{1+\eta} = O\left(T^{3\xi(1+\eta)}\right)$ and, since ξ < 1/6, we can substitute U_{Tt} for w_{Tt} in the proof of Theorem 3.

Similarly for V_{Tt}, we have

$$E\left|V_{Tt}\right|^{1+\eta} = E\left|\frac{T^{-3\xi}\varepsilon_{t-1}^2\varepsilon_{t-2}^2}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)^2}\right|^{1+\eta} = O\left(T^{3\xi(1+\eta)}\right)$$

and we can substitute V_{Tt} for w_{Tt} in the proof of Theorem 3.

Thus

$$\frac{c_0}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}^2}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)}\right) = c_0\omega_{T\bar{\gamma}\bar{\gamma}} + o_p(1).$$

(iv) Similarly for term (iv), we have

$$E\left|\frac{T^{-\xi}\varepsilon_{t-1}\varepsilon_{t-2}}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)}\right| = O\left(T^{3\xi}\right),$$

and substitution into the proof of Theorem 3 yields

$$\frac{c_0^2}{T^{3/2}}\sum_{t=1}^{T}\left(\frac{\varepsilon_{t-1}\varepsilon_{t-2}}{\left(c_0/\sqrt{T}+\varepsilon_{t-1}^2\right)\left(c_0/\sqrt{T}+\varepsilon_{t-2}^2\right)\left(\bar{c}/\sqrt{T}+\Delta y_{t-1}^2\right)}\right) \xrightarrow{L_1} 0\,.$$

Thus, for the numerator of $t_{\bar{\gamma}}$, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_t\,\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2} - \left(\omega_{T\sigma\bar{\gamma}}z - c_0\omega_{T\gamma\bar{\gamma}}\right) = o_p(1)$$

where $z \sim N(0, 1)$.

Consider the denominator of $t_{\bar{\gamma}}$. Now

$$\hat{u}_t = \Delta y_t - \hat{\phi}\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2},$$

where

$$\hat{\phi} = \frac{\frac{1}{T}\sum_{t=1}^{T} \Delta y_t \frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}}{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right)^2}.$$

Since

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Delta y_t \frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2} - \left(\omega_{T\sigma\bar{\gamma}} z - c_0 \omega_{T\gamma\bar{\gamma}}\right) = o_p(1)$$

from above, the numerator of $\hat{\phi}$ converges to zero. Then replacing ε_t with Δy_t in the proof of Theorem 3, we obtain that the denominator is asymptotically equivalent to a non-zero constant. Thus $\hat{\phi} \xrightarrow{L_1} 0$.

Thus we have, for the denominator of $t_{\bar{\gamma}}$,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{u}_t \Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right)^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t \Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right)^2 + o_p(1) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t \varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2} - \frac{c_0 \varepsilon_t \varepsilon_{t-2}}{\sqrt{T}\left(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2\right)\left(c_0/\sqrt{T} + \varepsilon_{t-2}^2\right)}\right)^2 + o_p(1) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t \varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right)^2 + o_p(1) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_t \varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2}\right)^2 + o_p(1) \\
&= \omega_{T\sigma\bar{\gamma}}^2 + o_p(1)
\end{aligned}$$

from (i) above. This leaves

$$t_{\bar{\gamma}} - \left(z - \frac{c_0 \omega_{T\gamma\bar{\gamma}}}{\omega_{T\sigma\bar{\gamma}}}\right) = o_p(1)$$

and the result follows. *Q.E.D.*

PROOF OF COROLLARY 5: We have

$$t_{\bar{\gamma}} = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varDelta y_t \sum_{i=1}^{t} \bar{\alpha}^{i-1} \frac{\varDelta y_{t-i}}{\bar{c}/\sqrt{T} + \varDelta y_{t-i}^2}}{\left(\frac{1}{T}\sum_{t=1}^{T}\left(\hat{u}_t \sum_{i=1}^{t} \bar{\alpha}^{i-1} \frac{\varDelta y_{t-i}}{\bar{c}/\sqrt{T} + \varDelta y_{t-i}^2}\right)^2\right)^{1/2}}$$

$$= \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varDelta y_t \bar{C}(L)\left(\frac{\varDelta y_{t-1}}{\bar{c}/\sqrt{T} + \varDelta y_{t-1}^2}\right)}{\left(\frac{1}{T}\sum_{t=1}^{T}\left(\hat{u}_t \bar{C}(L)\left(\frac{\varDelta y_{t-1}}{\bar{c}/\sqrt{T} + \varDelta y_{t-1}^2}\right)\right)^2\right)^{1/2}}$$

where $\bar{C}(L) = 1 + \bar{\alpha}L + \bar{\alpha}^2 L^2 + ...$, and L denotes the lag operator. Define

$$\bar{x}_{Tt-1} = \bar{C}(L)\left(\frac{\varepsilon_{t-1}}{\bar{c}/\sqrt{T} + \varepsilon_{t-1}^2}\right),\ x_{Tt-1} = C_0(L)\left(\frac{\varepsilon_{t-1}}{c_0/\sqrt{T} + \varepsilon_{t-1}^2}\right),$$

where $C_0(L) = 1 + \alpha_0 L + \alpha_0^2 L^2 + ...$, and note that the process in (7) can be written as

$$\Delta y_t = \varepsilon_t + \frac{c_0}{\sqrt{T}}(\alpha_0 - 1)x_{Tt-1}.$$

The only change from the case in Theorem 4 is that, since x_{Tt} depends on the infinite past of ε_t, we cannot directly map the mixing properties of ε_t to x_{Tt}. However we can utilize the concept of near epoch dependence (NED). From Davidson (1994, Definition 17.1), x_{Tt} is L_p-NED if

$$\left(E\left|x_{Tt} - E\left(x_{Tt}\left|\Im_{t-j}^{t+j}\right.\right)\right|^p\right)^{1/p} \le d_{Tt}\nu_j$$

where $p > 0$, $\nu_j \to 0$ and $\{d_t\}$ is a positive sequence of constants. We have

$$E\left|x_{Tt} - E\left(x_{Tt}\left|\Im_{t-j}^{t+j}\right.\right)\right| = E\left|\sum_{i=0}^{\infty}\alpha^i \frac{\varepsilon_{t-i}}{c/\sqrt{T} + \varepsilon_{t-i}^2} - E\left(\sum_{i=0}^{\infty}\alpha^i \frac{\varepsilon_{t-i}}{c/\sqrt{T} + \varepsilon_{t-i}^2}\left|\Im_{t-j}^{t+j}\right.\right)\right|$$

$$= E\left|\sum_{i=j+1}^{\infty}\alpha^i \left(\frac{\varepsilon_{t-i}}{c/\sqrt{T} + \varepsilon_{t-i}^2} - E\left(\frac{\varepsilon_{t-i}}{c/\sqrt{T} + \varepsilon_{t-i}^2}\left|\Im_{t-j}^{t+j}\right.\right)\right)\right|$$

$$\le \sum_{i=j+1}^{\infty}\alpha^i E\left|\frac{\varepsilon_{t-i}}{c/\sqrt{T} + \varepsilon_{t-i}^2} - E\left(\frac{\varepsilon_{t-i}}{c/\sqrt{T} + \varepsilon_{t-i}^2}\left|\Im_{t-j}^{t+j}\right.\right)\right|$$

$$\leq 2E\left|\frac{\varepsilon_t}{c/\sqrt{T}+\varepsilon_t^2}\right|\sum_{i=j+1}^{\infty}\alpha^i$$

Let $v_j = \sum_{i=j+1}^{\infty}\alpha^i \to 0$ and

$$d_{Tt} = 2E\left|\frac{\varepsilon_t}{c/\sqrt{T}+\varepsilon_t^2}\right| = O\left(T^{\xi}\right)$$

where $0 < \xi < 1/6$. Given this, and the law of large numbers in de Jong (1995, Theorem 3) and the central limit theorem in de Jong (1997, Theorem 2), the results go through in identical fashion to those in the proof of Theorem 4. We present an outline below.

For the numerator of $t_{\bar{\gamma}}$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_t\overline{C}(L)\left(\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)$$

$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\varepsilon_t+\frac{c_0}{\sqrt{T}}(\alpha_0-1)x_{Tt-1}\right)\overline{C}(L)\left(\frac{\varepsilon_{t-1}+\frac{c_0}{\sqrt{T}}(\alpha_0-1)x_{Tt-2}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)$$

$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\overline{C}(L)\left(\frac{\varepsilon_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)+\frac{c_0}{T}(\alpha_0-1)\sum_{t=1}^{T}x_{Tt-1}\overline{C}(L)\left(\frac{\varepsilon_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)$$

$$+\frac{c_0}{T}(\alpha_0-1)\sum_{t=1}^{T}\varepsilon_t\overline{C}(L)\left(\frac{x_{Tt-2}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)$$

$$+\frac{c_0^2}{T^{3/2}}(\alpha_0-1)^2\sum_{t=1}^{T}x_{Tt-1}\overline{C}(L)\left(\frac{x_{Tt-2}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right)$$

$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\bar{x}_{Tt-1}+\frac{c_0}{T}(\alpha_0-1)\sum_{t=1}^{T}x_{Tt-1}\bar{x}_{Tt-1}+o_p(1)$$

$$=\omega_{T\bar{\alpha}\bar{\gamma}}z+c_0(\alpha_0-1)\omega_{T\bar{\alpha}\bar{\gamma}\bar{\gamma}}+o_p(1)$$

where z ~ N(0, 1).

For the denominator, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{u}_t\sum_{i=1}^{t}\bar{\alpha}^{i-1}\frac{\Delta y_{t-i}}{\bar{c}/\sqrt{T}+\Delta y_{t-i}^2}\right)^2=\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t\sum_{i=1}^{t}\bar{\alpha}^{i-1}\frac{\varepsilon_{t-i}}{\bar{c}/\sqrt{T}+\varepsilon_{t-i}^2}\right)^2+o_p(1)$$

$$= \omega^2_{T\bar{\alpha}\bar{\gamma}} + o_p(1)$$

Thus

$$t_{\bar{\gamma}} = c_0(\alpha_0 - 1)\frac{\omega_{T\bar{\alpha}\bar{\gamma}\bar{\gamma}}}{\omega_{T\bar{\alpha}\bar{\gamma}}} + z$$

and the result follows. *Q.E.D.*

PROOF OF THEOREM 6: Gaussianity of $L(y^T, \varphi)$ and compactness of Ψ provide sufficient regularity for the existence of the QMLE (White (1994), Theorem 2.12).

We have

$$T^{-1}L(y^T, \varphi) = -\frac{1}{2T\sigma^2}\sum_{t=1}^{T}(\Delta y_t - \alpha \Delta y_{t-1} + \theta_{t-1}\varepsilon_{t-1})^2 - \frac{1}{2}\log(2\pi\sigma^2)$$

$$= T^{-1}\sum_{t=1}^{T} \log \; f(y^t, \varphi).$$

To show consistency, we utilize the uniform law of large numbers (ULLN) for stationary ergodic processes of Ranga Rao (1962, reprinted as Theorem A.2.2 in White (1994)).

To invoke this ULLN, we need a dominating function $D_t = D(y^t)$ measurable-$\Im_t$ such that $\left|\log \; f(y^t, \varphi)\right| \leq D(y^t) \;\forall\; \varphi \in \Psi$ and $E(D_t) < \infty$. From the c_r inequality (White (1994), Proposition 3.8) we have,

$$(\Delta y_t - \alpha \Delta y_{t-1} + \theta_{t-1}\varepsilon_{t-1})^2 \leq 4\Delta y_t^2 + 4\Delta y_{t-1}^2 + 2\varepsilon_{t-1}^2$$

since $0 \leq \alpha < 1$ and $0 \leq \theta_{t-1} \leq 1$. Thus, apart from a constant, a dominating function is

$$D_t = 4\Delta y_t^2 + 4\Delta y_{t-1}^2 + 2\varepsilon_{t-1}^2,$$

which is finite in expectation if Δy_t has finite second moments.

$$E(\Delta y_t^2) = E\left[\left((\alpha - 1)\sum_{i=1}^{t}\alpha^{i-1}(1 - q_{t-i})\varepsilon_{t-i} + \varepsilon_t\right)^2\right]$$

$$= E\left[\left((\alpha - 1)\sum_{i=1}^{t}\alpha^{i-1}(1 - q_{t-i})\varepsilon_{t-i}\right)^2\right] + \sigma_0^2$$

$$= (1-\alpha)^2 E\left[\sum_{i=1}^{t}\left(\alpha^{i-1}(1-q_{t-i})\varepsilon_{t-i}\right)^2\right] + \sigma_0^2$$

since $\{\varepsilon_t, \Im_t\}$ and $\{q_t\varepsilon_t, \Im_t\}$ are martingale difference sequences by assumption. Now

$$(1-\alpha)^2 E\left[\sum_{i=1}^{t}\left(\alpha^{i-1}(1-q_{t-i})\varepsilon_{t-i}\right)^2\right] < (1-\alpha)^2 E\left[\sum_{i=1}^{t}\left(\alpha^{i-1}\varepsilon_{t-i}\right)^2\right] < \infty \quad \forall \varphi \in \Psi.$$

since $0 \leq \alpha < 1$.

Since ε_t is stationary and ergodic and because $L(y^T, \varphi)$ is a measurable function of ε_t, $L(y^T, \varphi)$ is also stationary and ergodic (see White (1984, Theorem 3.35)). Thus, the ULLN holds, i.e. $T^{-1}L(y^T, \varphi) - E\left(T^{-1}L(y^T, \varphi)\right) \xrightarrow{a.s.} 0$ uniformly on Ψ. It then follows from uniqueness of φ_0 and Theorem 3.4 of White (1994), that $\hat{\varphi} - \varphi_0 \xrightarrow{a.s.} 0$. *Q.E.D.*

The following Lemma will prove useful in completing the proof of Theorem 7.

LEMMA A2: *Suppose that ε_t, t = 0, 1, …, T, is a random variable with a continuous density function and variance $\sigma_0^2 > 0$. Then $|s_t| < \infty$ wp 1 and $|\partial s_t / \partial \varphi_{12}| < \infty$ wp 1, where $s_t = (w_t \; v_t)'$ as defined in Equations (15)–(17), $\varphi_{12} = (\gamma \;\; \alpha)$ and $q_t(\gamma) = \varepsilon_t^2 / (\gamma + \varepsilon_t^2)$.*

PROOF OF LEMMA A2: We have

$$w_t = b_{t-1} w_{t-1} - (1-\alpha)\frac{\partial q_{t-1}}{\partial \gamma}\varepsilon_{t-1},$$

$$v_t = b_{t-1} v_{t-1} + q_{t-1}\varepsilon_{t-1} - \Delta y_{t-1},$$

$$b_{t-1} = 1 - (1-\alpha)(1+\eta_{t-1})q_{t-1},$$

and $w_0 = v_0 = 0$ wp 1. Consider

$$|w_t| \leq |b_{t-1} w_{t-1}| + (1-\alpha)\left|\frac{\varepsilon_{t-1}^3}{\left(\gamma + \varepsilon_{t-1}^2\right)^2}\right|$$

Now

$$\left|\frac{\varepsilon_{t-1}^3}{\left(\gamma+\varepsilon_{t-1}^2\right)^2}\right| \le \frac{3\sqrt{3}}{16\sqrt{\gamma}} \le \frac{1}{\sqrt{\gamma}}$$

and $\left|b_{t-1}\right| < 1$ wp 1, which implies

$$\left|w_t\right| < \bar{b}\left|w_{t-1}\right| + (1-\alpha)\gamma^{-1/2} \quad \text{wp 1,}$$

where $0 < \bar{b} < 1$. Then, since $w_0 = 0$ wp 1, we have

$$\left|w_t\right| < \frac{(1-\alpha)}{\left(1-\bar{b}\right)\sqrt{\gamma}} < \infty \quad \forall\varphi \in \Psi \ .$$

Also, since

$$\left|\frac{\varepsilon_{t-1}^3}{\left(\gamma+\varepsilon_{t-1}^2\right)} - \Delta y_{t-1}\right| \le \left|\frac{\gamma\varepsilon_{t-1}}{\gamma+\varepsilon_{t-1}^2}\right| + (1-\alpha)\gamma\left|\sum_{i=1}^{t}\alpha^{i-1}\frac{\gamma\varepsilon_{t-1-i}}{\gamma+\varepsilon_{t-1-i}^2}\right| \le \frac{1}{\sqrt{\gamma}} \ ,$$

we have

$$\left|v_t\right| \le \bar{b}\left|v_{t-1}\right| + \gamma^{-1/2} \quad \text{wp 1,}$$

which implies

$$\left|v_t\right| < \frac{1}{\left(1-\bar{b}\right)\sqrt{\gamma}} < \infty \quad \forall\varphi \in \Psi \ ,$$

using $v_0 = 0$ wp 1. Thus $\left|s_t\right| < \infty$ wp 1.

Now $\partial s_t / \partial \varphi_{12}$ is a symmetric 2×2 matrix with the following components:

$$\frac{\partial w_t}{\partial \gamma} = b_{t-1}\frac{\partial w_{t-1}}{\partial \gamma} + \frac{\partial b_{t-1}}{\partial \gamma}w_{t-1} - (1-\alpha)\frac{2\varepsilon_{t-1}^3}{\left(\gamma+\varepsilon_{t-1}^2\right)^3} \ ,$$

$$\frac{\partial w_t}{\partial \alpha} = \frac{\partial v_t}{\partial \gamma} = b_{t-1}\frac{\partial v_{t-1}}{\partial \gamma} + \frac{\partial b_{t-1}}{\partial \gamma}v_{t-1} - \frac{\varepsilon_{t-1}^3}{\left(\gamma+\varepsilon_{t-1}^2\right)^2} \ ,$$

$$\frac{\partial v_t}{\partial \alpha} = b_{t-1}\frac{\partial v_{t-1}}{\partial \alpha} + \frac{\partial b_{t-1}}{\partial \alpha}v_{t-1} \ ,$$

where

$$\frac{\partial b_{t-1}}{\partial \gamma} = (1-\alpha)\frac{\left(3\gamma-\varepsilon_{t-1}^2\right)\varepsilon_{t-1}^2}{\left(\gamma+\varepsilon_{t-1}^2\right)^3} \ ,$$

$$\frac{\partial b_{t-1}}{\partial \alpha} = \frac{\left(3\gamma + \varepsilon_{t-1}^2\right)\varepsilon_{t-1}^2}{\left(\gamma + \varepsilon_{t-1}^2\right)^2}.$$

Using

$$\left|\frac{\partial b_{t-1}}{\partial \gamma}\right| \le \frac{5\sqrt{13}-17}{\left(5-\sqrt{13}\right)^3 \gamma} < \frac{1}{\gamma} \quad \text{and} \quad \left|\frac{2\varepsilon_{t-1}^3}{\left(\gamma + \varepsilon_{t-1}^2\right)^3}\right| \le \frac{1}{4\gamma^{3/2}}$$

and $|b_{t-1}| < 1$ wp 1 yields

$$\left|\frac{\partial w_t}{\partial \gamma}\right| \le \left|b_{t-1}\frac{\partial w_{t-1}}{\partial \gamma}\right| + \left|\frac{\partial b_{t-1}}{\partial \gamma} w_{t-1}\right| + (1-\alpha)\left|\frac{2\varepsilon_{t-1}^3}{\left(\gamma + \varepsilon_{t-1}^2\right)^3}\right|$$

$$< \bar{b}\left|\frac{\partial w_{t-1}}{\partial \gamma}\right| + |w_{t-1}|(1-\alpha)\gamma^{-1} + (1-\alpha)\frac{1}{4\gamma^{3/2}}$$

where $0 < \bar{b} < 1$. From above, we have $|w_{t-1}| < \infty$ wp 1. Then $w_0 = 0$ wp 1 implies $|\partial w_t / \partial \gamma| < \infty$ wp 1. Similarly, we have

$$\left|\frac{\partial w_t}{\partial \alpha}\right| = \left|\frac{\partial v_t}{\partial \gamma}\right| < \bar{b}\left|\frac{\partial v_{t-1}}{\partial \gamma}\right| + \gamma^{-1}|v_{t-1}| + \frac{3\sqrt{3}}{16\sqrt{\gamma}} < \infty \quad \text{wp 1},$$

and

$$\left|\frac{\partial v_t}{\partial \alpha}\right| < \bar{b}\left|\frac{\partial v_{t-1}}{\partial \alpha}\right| + \frac{9}{8}|v_{t-1}| < \infty \quad \text{wp 1.}$$

Q.E.D.

PROOF OF THEOREM 7: Consider a first order mean value expansion of the first order condition around φ_0. Define $\nabla_\varphi^2 L(y^T,\varphi)$ as the hessian. The mean value expansion yields

$$\nabla_\varphi L\left(y^T, \hat{\varphi}\right) = 0 = \nabla_\varphi L\left(y^T, \varphi_0\right) + \nabla_\varphi^2 \ddot{L}\left(y^T, \varphi\right)(\hat{\varphi} - \varphi_0)$$

where $\nabla_\varphi \ddot{L}\left(y^T, \varphi\right)$ has each row evaluated at a mean value (possibly different for each row) lying between $\hat{\varphi}$ and φ_0. We can rearrange to get

$$\sqrt{T}(\hat{\varphi} - \varphi_0) = H^{-1}T^{-1/2}\nabla_\varphi L(y^T, \varphi_0) + \left(\left(T^{-1}\nabla_\varphi^2 \ddot{L}(y^T, \varphi)\right)^{-1} - H^{-1}\right)\nabla_\varphi L(y^T, \varphi_0)$$

$$= H^{-1}T^{-1/2}\nabla_\varphi L(y^T, \varphi_0) + o_p(1)$$

To verify this, we need to show that the hessian converges to H, which is nonsingular by assumption.

We partition the scores vector $T^{-1/2}\nabla_\varphi L(y^T,\varphi)$ as

$$T^{-1/2}\nabla_\varphi L\left(y^T,\varphi\right)=\begin{bmatrix} -\sigma^{-2}T^{-1/2}\sum_{t=1}^{T} s_t\varepsilon_t \\ \sigma^{-3}T^{-1/2}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right)\end{bmatrix}$$

where $s_t = (w_t, v_t)'$ as defined in (15) – (17). Similarly, the hessian is partitioned into the following matrix:

$$\nabla_\varphi^2 L\left(y^T,\varphi\right)=\begin{bmatrix}\tilde{H}_1 & \tilde{H}_3\\ \tilde{H}_3' & \tilde{H}_2\end{bmatrix}.$$

and consider each of the components $\tilde{H}_1, \tilde{H}_2$ and $\tilde{H}_3$ separately.

Now

$$\tilde{H}_1 = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(s_t s_t' + \varepsilon_t\frac{\partial s_t}{\partial\varphi_{12}}\right)$$

Since s_t is a measurable function of ε_t, s_t and $\partial s_t/\partial\varphi_{12}$ are stationary and ergodic (see White (1984, Theorem 3.35)). Then, given that $\left|s_t s_t'\right| < \infty$ wp 1 from Lemma A2, there exists a dominating function, D_t measurable-$\Im_t$, such that $\left|s_t s_t'\right| \le D\left(y^t\right)$ $\forall$ $\varphi \in \Psi$ and $E(D_t) < \infty$. Then from the uniform law of large numbers (ULLN) for stationary ergodic processes of Ranga Rao (1962, reprinted as Theorem A.2.2 in White (1994)), we have

$$T^{-1}\tilde{H}_1 - E\left(T^{-1}\tilde{H}_1\right)\xrightarrow{a.s.}0 \quad \text{uniformly on } \Psi,$$

where

$$E\left(T^{-1}\tilde{H}_1\right) = -\frac{1}{T\sigma^2}\sum_{t=1}^{T}E\left(s_t s_t'\right)$$

since $\{\varepsilon_t, \Im_t\}$ is a martingale difference sequence.

Consider

$$\tilde{H}_2 = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(3\varepsilon_t^2/\sigma^2 - 1\right).$$

Now, since $\sigma^2 > 0$, we can define $D_t = 3k^2\varepsilon_t^2 - k$ for some $k < \infty$ and use the Ranga Rao ULLN to obtain

$$T^{-1}\tilde{H}_2 - \frac{1}{T\sigma^2}\sum_{t=1}^{T}E\left(3\varepsilon_t^2/\sigma^2 - 1\right)\xrightarrow{a.s.}0 \quad \text{uniformly on } \Psi,$$

where

$$\frac{1}{T\sigma^2}\sum_{t=1}^{T} E\left(3\varepsilon_t^2/\sigma^2 - 1\right) = 2/\sigma^2 \ .$$

Consider

$$\tilde{H}_3 = \frac{2}{\sigma^3}\sum_{t=1}^{T} s_t\varepsilon_t \ .$$

Since $\sigma^2 > 0$ and $|s_t| < \infty$ wp 1, we can define $D_t = k|\varepsilon_t|$ for some $k < \infty$ and use the Ranga Rao ULLN to obtain

$$T^{-1}\tilde{H}_3 - \frac{2}{T\sigma^3}\sum_{t=1}^{T} E\left(s_t\varepsilon_t\right) \xrightarrow{a.s.} 0 \quad \text{uniformly on } \Psi.$$

From the law of iterated expectations, we have

$$\frac{2}{T\sigma^3}\sum_{t=1}^{T} E\left(s_t E\left(\varepsilon_t | \Im_{t-1}\right)\right) = 0 \ .$$

Thus, $T^{-1}\nabla_\varphi^2 L\left(y^T,\varphi\right) - E\left(T^{-1}\nabla_\varphi^2 L\left(y^T,\varphi\right)\right) \xrightarrow{a.s.} 0$ uniformly on Ψ, and we must now verify that $\left(T^{-1}\nabla_\varphi^2 \ddot{L}\left(y^T,\varphi\right) - H\right) \xrightarrow{a.s.} 0$. Consider

$$\left(T^{-1}\nabla_\varphi^2 L\left(y^T,\hat{\varphi}\right) - H\right) = \left(T^{-1}\nabla_\varphi^2 L\left(y^T,\hat{\varphi}\right) - E\left(T^{-1}\nabla_\varphi^2 L\left(y^T,\hat{\varphi}\right)\right)\right) + \left(E\left(T^{-1}\nabla_\varphi^2 L\left(y^T,\hat{\varphi}\right)\right) - H\right).$$

Now, the first term goes to zero from the above results on uniform convergence of the hessian to its expected value. Since $T^{-1}\nabla_\varphi^2 L\left(y^T,\varphi\right)$ is a continuous function of φ, consistency of $\hat{\varphi}$ for φ_0 ensures that the second term goes to zero almost surely (see White (1984, Proposition 2.11)). Thus

$$\left(T^{-1}\nabla_\varphi^2 L\left(y^T,\hat{\varphi}\right) - H\right) \xrightarrow{a.s.} 0$$

$$\Rightarrow \quad \left(T^{-1}\nabla_\varphi^2 \ddot{L}\left(y^T,\varphi\right) - H\right) \xrightarrow{a.s.} 0 \ ,$$

since the arguments of $\ddot{L}\left(y^T,\varphi\right)$ always lie between $\hat{\varphi}$ and φ_0.

Now consider

$$T^{-1/2}\nabla_\varphi L\left(y^T,\varphi_0\right) = \begin{bmatrix} -\sigma_0^{-2}T^{-1/2}\sum_{t=1}^{T} s_t\varepsilon_t \\ \sigma_0^{-3}T^{-1/2}\sum_{t=1}^{T}\left(\varepsilon_t^2 - \sigma_0^2\right) \end{bmatrix} \equiv T^{-1/2}\sum_{t=1}^{T} Z_t \ .$$

Since $T^{-1/2}\nabla_{\varphi}L(y^T,\varphi_0)$ is a measurable function of ε_t, which is stationary and ergodic, $T^{-1/2}\nabla_{\varphi}L(y^T,\varphi_0)$ is also stationary and ergodic (see White (1984, Theorem 3.35)). Thus we use Theorem 5.15 of White (1984), to show asymptotic normality. This requires $E(Z_0|\Im_{-m})\xrightarrow{q.m.}0$ as $m\to\infty$, $\sum_{j=0}^{\infty}(\mathrm{var}\,\Re_{0j})^{1/2}<\infty$ where $\Re_{0j}\equiv E(Z_0|\Im_{-j})-E(Z_0|\Im_{-j-1})$.

Since $\{s_t\varepsilon_t,\Im_t\}$ is a martingale difference sequence with finite variance, these conditions follow trivially for the first term in $\nabla_{\varphi}L(y^T,\varphi_0)$. For the second term, they hold by assumption. Thus we have

$$V^{-1/2}T^{-1/2}\nabla_{\varphi}L(y^T,\varphi_0)\xrightarrow{d}N(0,1),$$

and the result follows. *Q.E.D.*

REFERENCES

ANDREWS, D.W.K. (1993): "Tests for Parameter Instability and Structural Change with an Unknown Change Point," *Econometrica*, 61, 821-856.

ANDREWS, D.W.K., I. LEE, AND W. PLOBERGER (1996): "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics*, 70, 9-38.

CERCHI, M., AND A. HAVENNER (1988): "Cointegration and Stock Prices: The Random Walk on Wall Street Revisited," *Journal of Economic Dynamics and Control*, 12, 333-346.

CHELLEY-STEELEY, P.L., AND E.J. PENTECOST (1994): "Stock Market Efficiency, the Small Firm Effect and Cointegration," *Applied Financial Economics*, 4, 405-411.

CHOW, G.C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591-605.

CHRISTIANO, L.J. (1992): "Searching for a Break in GDP," *Journal of Business and Economic Statistics*, 10, 237-249.

COOK, R.D. AND S. WEISBURG (1990): "Confidence Curves in Nonlinear Regression," *Journal of the American Statistical Association*, 85, 544-551.

DAVIDSON, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, New York: Oxford University Press.

DAVIES, R.B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247-254.

DE JONG, R.M. (1995): "Laws of Large Numbers for Dependent Heterogeneous Processes," *Econometric Theory*, 11, 347-358.

DE JONG, R.M. (1997): "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory*, 13, 353-367.

GRANGER, C.W.J. (1986): "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics*, 48, 213-228.

GRANGER, C.W.J. AND A. ANDERSEN (1978): "On the Invertibility of Time Series Models," *Stochastic Processes and their Applications,* 8, 87-92.

GRANGER, C.W.J., AND N.R. SWANSON (1997): "An Introduction to Stochastic Unit Root Processes," *Journal of Econometrics*, 80, 35-62.

HAMILTON, J.D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica,* 57, 357-384.

HANSEN, B.E. (1992): "Testing for Parameter Instability in Regressions with I(1) Process," *Journal of Business and Economic Statistics*, 10, 321-335.

——— (1996): "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413-430.

HENDRY, D.F., AND M.P. CLEMENTS (1996): "Intercept Corrections and Structural Change," *Journal of Applied Econometrics*, 11, 475-494.

PHILLIPS, P.C.B., AND V. SOLO (1992): "Asymptotics for Linear Processes", *Annals of Statistics*, 20, 971-1001.

RANGA RAO, R. (1962): "Relations Between Weak and Uniform Convergence of Measures with Applications," *Annals of Mathematical Statistics*, 33, 659-680.

SAIKKONEN, P., AND R. LUUKKONEN (1993): "Testing for a Moving Average Unit Root in Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 88, 596-601.

SCHOENBERG, R. (1997): "Constrained Maximum Likelihood," *Computational Economics*, 10, 251-266.

SHEPHARD, N. (1994): "Partial non-Gaussian State Space," *Biometrika,* 81, 115-131.

STENGOS, T., AND E. PANAS (1992): "Testing the Efficiency of the Athens Stock Exchange: Some Results from the Banking Sector," *Empirical Economics*, 17, 239-252.

TONG, H. (1983): *Threshold Models in Non-linear Time Series Analysis,* Lecture Notes in Statistics, 21, Berlin: Springer.

WHITE, H. (1984): *Asymptotic Theory for Econometricians*, New York: Academic Press.

———— (1994): *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press.

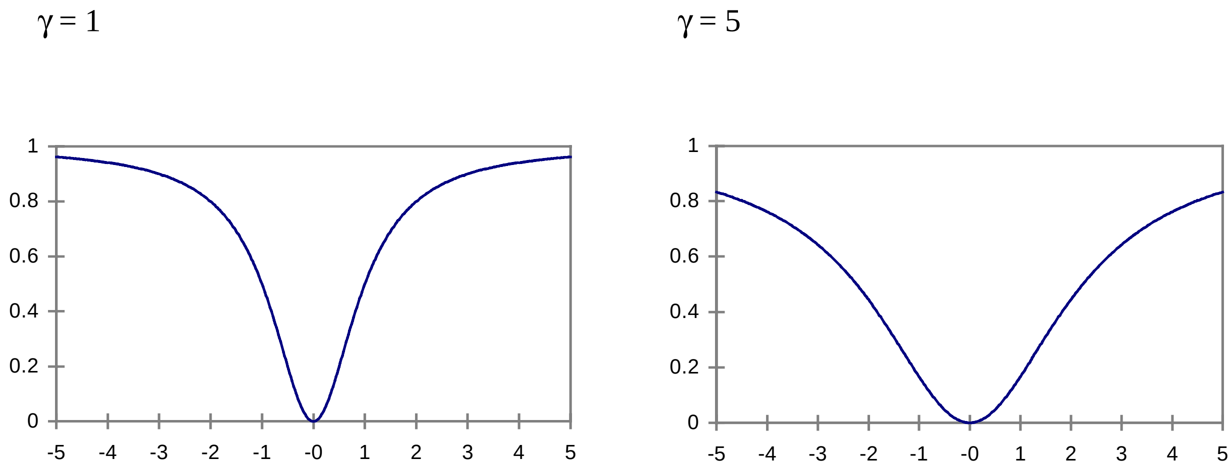
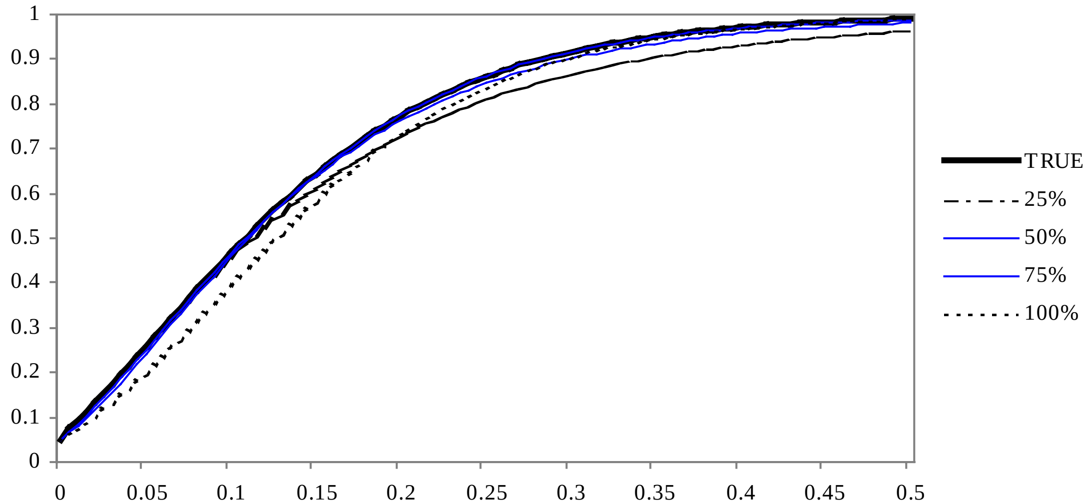


Figure 1 – $Q_T(\gamma)$ PLOTTED AGAINST ε_T



Notes: (i) Curves plot asymptotic power against γ^* for T=200.
(ii) The "TRUE" curve is optimal everywhere and the other curves are derived for fixed $\bar{c}$ such that power is optimized at the indicated level.

Figure 2 – ASYMPTOTIC POWER CURVES

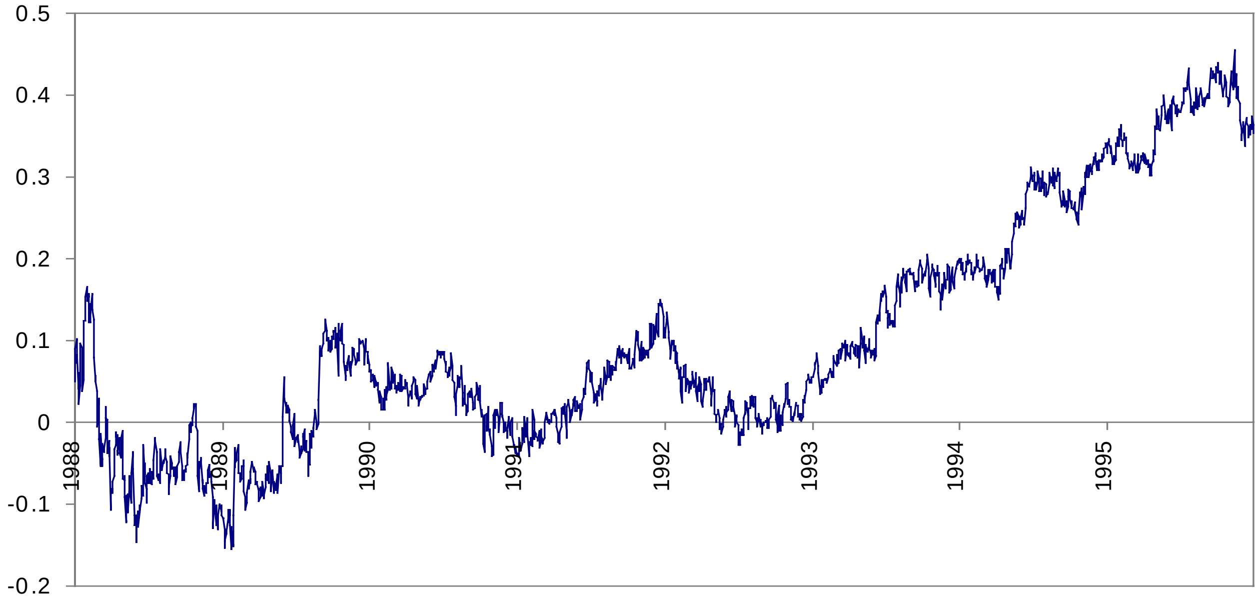


FIGURE 3: Log Share Price Ratio - Mobil/Texaco

TABLE I

POWER OF DF TEST AGAINST STOPBREAK

γ/σ^2	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Power	0.05	0.15	0.23	0.30	0.41	0.45	0.56	0.62	0.65	0.68	0.72

Monte Carlo experiment conducted at nominal size of 5% using 5000 repetitions on samples of 1000 observations.

TABLE II

CHOOSING LOCAL POINT ALTERNATIVE

γ^*	0.79	0.52	0.40	0.18	0.10	0.074	0.060	0.035	0.018	0.011
T	50	75	100	250	500	750	1000	2000	5000	10000

TABLE III

Modeling the Two Shock Process: Average Mean Square Forecast Errors

p	σ_u^2	Omniscient	Random Walk	Exp. Smoother	STOPBREAK	(Exp.Smoother − STOPBREAK)
0.01	2	1.02	2.02	1.16 (0.07)	1.16 (0.06)	0.00 (0.02)
0.05	2	1.10	2.10	1.38 (0.09)	1.38 (0.08)	0.00 (0.02)
0.1	2	1.20	2.20	1.58 (0.09)	1.58 (0.09)	0.00 (0.03)
0.01	5	1.05	2.05	1.26 (0.10)	1.22 (0.07)	0.04 (0.04)
0.05	5	1.25	2.25	1.66 (0.13)	1.59 (0.11)	0.07 (0.03)
0.1	5	1.50	2.50	2.03 (0.14)	1.95 (0.13)	0.08 (0.04)
0.01	10	1.10	2.10	1.38 (0.14)	1.28 (0.09)	0.09 (0.06)
0.05	10	1.50	2.50	2.02 (0.19)	1.86 (0.16)	0.15 (0.05)
0.1	10	2.00	3.00	2.66 (0.23)	2.49 (0.21)	0.16 (0.05)
0.01	20	1.20	2.20	1.57 (0.21)	1.39 (0.14)	0.18 (0.08)
0.05	20	2.00	3.00	2.64 (0.31)	2.38 (0.28)	0.26 (0.06)
0.1	20	3.00	4.00	3.79 (0.40)	3.52 (0.38)	0.26 (0.06)

TABLE IV

TESTING THE NULL OF A SIMPLE RANDOM WALK AGAINST STOPBREAK

	$\text{Sup}_{\alpha}\ t_{\bar{\gamma}}$	Approx. (p=5)	Approx. (p=10)	$\alpha = 0.8$
Coke/Pepsi	-2.98**	15.08**	21.94**	-2.88**
J&J/Merck	-4.02**	18.00**	28.65**	-4.01**
GM/Ford	-3.23**	17.91**	20.78**	-3.08**
Mobil/Texaco	-5.11**	22.55**	33.20**	-5.11**
ITT/Hilton	-2.12**	16.25**	19.22**	-1.96**
AT&T/MCI	-2.24**	8.93	13.47	-2.24**
McDD/Boeing	-0.03	5.29	11.62	0.28
IBM/Microsoft	-2.60**	7.37	13.46	-2.60**
Coors/Anheuser-Busch	-3.70**	13.63**	21.48**	-3.66**
Critical Values: 10%	-1.72	9.23	15.99	-1.28
Critical Values: 5%	-2.07	11.07	18.31	-1.65

Notes: (i) ** indicates significance at 5% and * significance at 10%

(ii) Critical values for the $\text{sup}_{\alpha}\ t_{\bar{\gamma}}$ test were simulated using 1000 repetitions on 1000 observations and $\bar{\alpha} \in [0,0.9]$.

TABLE V

QMLE ESTIMATION

	Coke/Pepsico	Mobil/Texaco	J&J/Merck	Coors/Anheuser
$\hat{\alpha}$	0.711	0.746	0.739	0.830
95% C.I.	(0.476, 0.966)	(0.623, 0.842)	(0.495, 0.970)	(0.681, 0.925)
Usual s.e.	(0.089)	(0.037)	(0.088)	(0.041)
Het. Cons. s.e.	(0.172)	(0.044)	(0.200)	(0.060)
$\hat{\gamma}$ $(\times 10^3)$	0.065	0.101	0.090	0.308
95% C.I.	(0.019, 0.323)	(0.044, 0.195)	(0.025, 1.210)	(0.088, 0.782)
Usual s.e.	(0.023)	(0.025)	(0.038)	(0.114)
Het. Cons. s.e.	(0.040)	(0.038)	(0.082)	(0.184)
$\hat{\gamma}/\hat{\sigma}^2$	0.384	0.766	0.517	0.617
$\hat{\sigma}^2$	0.169×10^{-3}	0.131×10^{-3}	0.175×10^{-3}	0.500×10^{-3}
$T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2$	0.170×10^{-3}	0.133×10^{-3}	0.177×10^{-3}	0.504×10^{-3}

Notes: (i) Optimization performed using BFGS algorithm in GAUSS.

TABLE VI

TRADING STRATEGY PROFITS

	Coke/ Pepsico	Mobil/ Texaco	J&J/ Merck	Coors/ An-Busch
STOPBREAK:				
Average Annual Wealth	0.188	0.242	0.471	0.365
	(0.296)	(0.180)	(0.229)	(0.447)
'Sharpe Ratio'	0.634	1.342	2.057	0.818
Ave. Days Between Changes	2.62	1.96	1.96	2.36
Exponential Smoother:				
Average Annual Wealth	0.162	0.214	0.302	0.301
	(0.296)	(0.181)	(0.230)	(0.447)
'Sharpe Ratio'	0.568	1.185	1.314	0.674
Ave. Days Between Changes	2.16	1.94	1.99	2.05
20 Day Moving Average:				
Average Annual Wealth	0.307	0.121	0.442	0.406
	(0.296)	(0.181)	(0.229)	(0.446)
'Sharpe Ratio'	1.039	0.671	1.927	0.909
Ave. Days Between Changes	1.42	1.29	1.32	3.91

[1] This work has benefited from useful conversations with Graham Elliott, Clive Granger and Hal White. All remaining errors are our own.

[2] The abbreviation "wp 1" is used throughout the paper to abbreviate "with probability one."

[3] Throughout the paper, the notation $\tilde{x}_t$ will be used to indicate a realization on the random variable x_t.

[4] At this point there is no upper bound on λ_t. However, sufficient conditions for invertibility of the moving average representation of the process will be shown to suggest some logical bounds.

[5] Since the best linear representation of a STOPBREAK process is an integrated MA(1), the result in Theorem 2 could be interpreted as relating the size distortion of the DF test to γ, rather than giving the power of the test against γ.

[6] If ε_t is distributed independently, the value that we choose for $\bar{\alpha}$ has no relation to the optimal choice of $\bar{c}$ since changing $\bar{\alpha}$ serves only to scale the power against a non-zero c_0 value proportionally. Otherwise, we speculate that changing $\bar{\alpha}$ will have little effect on the optimal $\bar{c}$.

[7] Kullback-Leibler Information Criterion (see White (1994), Definition 2.2).

[8] The Sharpe Ratio is usually computed as the ratio of annual return to annual standard deviation. In this case, since there is no initial investment, we do not have a return but rather an accumulated wealth. Thus the use of the term "Sharpe Ratio" here is not strictly correct.

[9] It should be noted that this trading strategy requires regular transactions, as evidenced by the "average days between changes" in Table VI being around two in most cases. This means that the investor must reverse her position to go long in the other stock on average once every two days. For this reason transactions costs could potentially be high.